LACUS FORUM L

LANGUAGES WITHOUT BORDERS

ST. JEROME'S UNIVERSITY

UNIVERSITY OF WATERLOO

LACUS FORUM

Journal of the Linguistic Association of Canada and the United States

Editors of Volume L

Svetlana Kaminskaïa, Managing Editor
Dept. of French Studies/ Dép. d'études françaises
University of Waterloo / Université de Waterloo
200 University Ave. W., Waterloo, ON N2L 3G1

Jocelyn van Leeuwen
Dept. of French Studies/ Dép. d'études françaises
University of Waterloo / Université de Waterloo
200 University Ave. W., Waterloo, ON N2L 3G1

Editorial Assistant
Mariam Sidaoui
Dept. of French Studies/ Dép. d'études françaises
University of Waterloo / Université de Waterloo
200 University Ave. W., Waterloo, ON N2L 3G1

Editorial Help
Mona Ewais, James Mason, Grace Mercer
Dept. of French Studies/ Dép. d'études françaises
University of Waterloo / Université de Waterloo
200 University Ave. W., Waterloo, ON N2L 3G1

Renison Joseph Gonsalves, Production Editor
English Department & Linguistics Program
Brooklyn College (Emeritus)
2900 Bedford Ave.,
Brooklyn, NY 11210, U.S.A.

Reviewers of Volume L

Lilly Chen
Wladyslaw Cichocki
Grażyna Drzazga
Sheila Embleton
Christian Guilbault
Randy Harris
Liza Kopylova
Jacques Lamarche
Maciej Laskowski
Nicole Nolette
Troy Spier
Magda Stroińska
Jeff Tennant
Sarah Tsiang
Egor Tsedryk
Artur Urbaniak
Jocelyn van Leeuwen
Eric S. Wheeler
Joyce Zeng

LACUS Forum is published by the Linguistic Association of Canada and the United States.

Copyright © LACUS, 2025. All rights reserved.

Printed in U.S.A. ISSN 0195-377x

LACUS FORUM

Journal of the Linguistic Association of Canada and the United States

VOLUME 50 NUMBER 1, 2025

CONTENTS

Preface ii
Svetlana Kaminskaïa & Jocelyn van Leeuwen

1. *Presidential Address*
A linguist's adventures in the wonderland of translation 1
Magda Stroińska

2. The relationship between articulation rate and utterance length in Canadian French 13
Wladyslaw Cichocki, Svetlana Kaminskaïa & Luke Hagar

3. Cemeteries as semiotic landscapes: The construction of (im)migrant spaces in Canada and Germany 27
Grit Liebscher & Doris Stolberg

4. *Presidents' Prize Winner*
Some demolinguistic trends of the Francophone minority in Canada in the 21st century 40
Jocelyn van Leeuwen

5. *Presidents' Prize Winner*
Sauce Brune, Google Translate and ChatGPT: Translating Quebecois swear words 54
Liza Kopylova

PREFACE

The fiftieth annual forum of the Linguistic Association of Canada and the United States was hosted by the University of Waterloo and St. Jerome's University from July 17 to July 19, 2024, in Waterloo, Ontario. The local organizer (Svetlana Kaminskaïa, French Studies) thanks St. Jerome's University and the Department of French Studies for their generosity and help in welcoming participants from different parts of the world, both in person and online.

The forum's theme, "Languages Without Borders," invited papers that examine the role of language in shaping identities, influencing social dynamics, facilitating learning, creativity, and collaboration. The forum opened with a reception and an opening ceremony. Dean of Arts, Alexie Tcheuyap, and Vice President Academic and Dean of St. Jerome's University, Carol Ann MacGregor, cordially welcomed the participants.

Over the duration of the conference, attendees heard three plenary talks. They were presented by three professors from the University of Waterloo: Olga Vechtomova, Professor in the Department of Management Science and Engineering (Faculty of Engineering and Cheriton School of Computer Science); Nicole Nolette, Canada Research Chair in Minority Studies and Associate Professor in the Department of French Studies (Faculty of Arts); and Grit Liebscher, Professor of Applied Linguistics in the Department of Germanic and Slavic Studies (Faculty of Arts). We thank the plenary speakers for their contributions to the success of the conference.

Besides the keynote addresses, over 30 regular presentations were given at the conference. This volume includes 15 papers selected for publication among the submitted ones after review by at least two anonymous reviewers.

Magda Stroińska's presidential address offers a perspective of an applied linguist and translator on the field and process of translation. Using personal anecdotes and experiences, she illustrates the captivating and enriching nature of translation and invites linguists and specialists of cognitive study of language to offer greater focus to this discipline.

Wladyslaw Cichocki, Svetlana Kaminskaïa and Luke Hagar examine the relationship between articulation rate (AR) and interpausal intervals (IPIs) in three regional varieties of Canadian French. The researchers find that in all regions IPI length predicts AR but does not explain it, given differences observed in vowel durations in shorter and longer IPIs.

Grit Liebscher and Doris Stolberg examine cemeteries as semiotic landscapes. They study gravestone inscriptions and images in Canada and Germany to show how heritage languages and images participate in the construction of (im)migrant spaces and imagined communities.

Based on 2001-2021 census data, Jocelyn van Leeuwen, a winner of the Presidents' Prize for 2024, presents demolinguistic trends of the Francophone minority in Canada. The author takes into consideration dynamics related to mother tongue, home language, language shift, linguistic exogamy, and language transmission.

The article by another winner of the Presidents' Prize for 2024, Liza Kopylova, compares the results of translation and retranslation of *sacres* in a Quebecois play completed by Google Translate and ChatGPT. The author demonstrates that ChatGPT outperformed Google Translate in this task, but neither tool matched the frequency or intensity of profanities of the original text.

In her article, Olga Vechtomova examines how generative AI models can be used as a source of artistic inspiration. She compares two generative AI approaches and relies on concepts of randomness and predictability to demonstrate how AI-generated poetic lines can stimulate creativity.

Grażyna Drzazga and Magda Stroińska examine the change in Polish attitudes towards Ukrainian refugees as manifested in social media posts between 2022 and 2025. Besides observing a growth of negative attitudes, the authors note that other voices – those of support and empathy – give hope for healing and renewed relations.

Sheila Embleton and Eric Wheeler adopt a Multidimensional Scaling approach to explain dialect variation in Chinese and Mambila languages through their geography. The method, previously very successfully applied to English, Finnish and Romanian data brought "negative" results, thus inviting the researchers to further reflect on the definition of *geography* and *dialect variation*, and on the value of appropriate and sufficient data.

DeAndré A. Espree-Conaway's contribution examines the word class of body part terms in Tu'un Javi, an understudied language spoken in Mexico. The author demonstrates that these terms show a distribution different from their nominal origins and thus contribute to prepositional constructions.

Artur Urbaniak presents a study of attitudes toward feminatives in Polish public discourse from the Gen Z perspective. The author shows that while being open to these forms, participants are not using them as widely as expected and explains this by the complexity of morphological formation of some new feminatives and by their social perception.

Svetlana Kaminskaïa and Luke Hagar adopt a novel approach to examine the extent of rhythmic variation in a minority French variety. By examining two styles of speech and considering linguistic factors, the authors contribute new details to the description of Canadian French.

Jacques Lamarche presents an account of verb-noun conversion in English in a model of grammar where the lexical primitives of syntax are the phonological matrices of words identifying denotational content, and their grammatical values are derived from their position.

Maciej Laskowski's article presents an analysis of Canada's linguistic and cultural worldview as it appears in early 20th century memoirs written by Polish immigrants in Canada. These memoirs offer a valuable documentation of individual experiences and two countries' shared history.

Tatiana Pizarro analyzes sociodiscursive representations of women in the context of pension laws in Argentina. This paper examines legal and parliamentary discourse related to the recognition of housewives' retirement status and stresses the implications of discursive constructions in the participation and creation of social (in)equalities.

Joyce Zeng examines language brokery phenomenon among Chinese-English bilinguals by focusing on how they process idioms. The author found that structural properties of the idioms, language exposure, and translation experience of language brokers contribute to successful communication bridging languages and cultures.

We thank the authors for their contributions to this volume. We thank all the referees for their time and attention given to the papers. A special thank you to the undergraduate students at the Department of French studies – Mariam Sidaoui, Mona Ewais, James Mason, and Grace Mercer – for their assistance in formatting and proofreading the texts.

Svetlana Kaminskaïa and Jocelyn van Leeuwen, July 2025

LACUS Forum **50:1.1-12 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

A linguist's adventures in the wonderland of translation

Magda Stroińska
McMaster University

Abstract: The study of translation is considered a subfield of applied linguistics but not every translator is a linguist and not every linguist is interested in translation. My academic path started with German philology, then took me to the study of applied linguistics with a specialization in translation, to end with general linguistics and the cognitive science of language. Throughout all the stages and meanders of my career, translation remained present as an academic research interest, teaching subject, and an occasional source of income. This paper reviews several different aspects of translation as seen from the perspective of a linguist and emphasizes that its study is a rewarding research topic in the field of linguistics, offering unique insight into the way we process information in multiple codes and in the context of multiple cultures.

Keywords: translation, interpretation, teaching translation theory and practice, translating political discourse, court interpretation

Languages: German, English, Polish

1. Introduction

THE TITLE THAT I ORIGINALLY CONTEMPLATED for this presentation mentioned working in the world's oldest profession, but I later thought that it could easily be misconstrued. There are also many theories about what the oldest profession in the world is. While some may assume it is prostitution, this approach is relatively modern, often attributed to the literary work of Rudyard Kipling. Other professions have also made claims to be the oldest: spies, doctors, soldiers, and teachers. Here, I would like to talk about my own experiences with what I consider a very old human occupation, without necessarily trying to convince everyone that it is the oldest one: translation.

There is little agreement on how far into the past *professional* translation activity goes. Some say it was the Hebrew translation of the Bible in the 3rd century but the Rosetta Stone exercise in translation is much older, going back to the third millennium before the Christian era. A significant problem in this dating approach is that we, by necessity, only look at written translation, forgetting, it seems, that long before writing systems were developed, there must have been people translating through oral communication. Today, we distinguish oral translation, which we call *interpretation* and written translation, often considered as *translation proper*.

The patron saint of translators is St. Jerome of Stridon (c. 342–347 – 420 AD), a Christian priest,

theologian, and translator, best known for his translation of the Bible into Latin. As the version of Latin used in everyday communication was referred to as *common* (in Latin: *vulgar*), St. Jerome's version of the Bible was called *Vulgate*. His translation of the Old Testament was based on the Hebrew version, rather than the so-called Septuagint, the Greek translation done by 72 translators. In a 16th century painting by Albrecht Dürer, depicting St. Jerome in his study, the left hand of the saint rests on a human skull with the index finger pointing to the left temporal lobe, the brain centre for language comprehension and memory. What a curious coincidence! It was only in the mid-19th century that the left lateralization of language centres started to be researched.

I would like to share with you the multiple ways my personal and professional life brought me over the years in touch with various aspects of translation and interpretation and what insights these experiences gave me into the way we process languages.

2. A dream profession

My own adventures with translation started in my early teens with my admiration for one of my half-brothers, 25 years older than me. He, Andrzej Czcibor-Piotrowski, was a literary translator and a poet who worked primarily with Czech and Slovak, but also translated poetry from Russian, Ukrainian, and English. I looked up to him and could not imagine anything more fulfilling and romantic than working at home on a typewriter, bringing to life in one language texts that have been written in other languages and cultures. I was particularly proud of his Polish translation of T.S. Elliott's *The Waste land*. My brother studied Slavic philology at Warsaw University, specializing in Czech and Slovak languages. As a child, I did not fully understand that his knowledge of Russian, Ukrainian, and English came from the time he spent in a deportation camp in Siberia during WWII and then being transported, with many Polish orphans, to Scotland. I wanted to be a translator, just like my brother.

In high school, my Plan A was to study graphic design but when I failed to gain acceptance to the Academy of Fine Arts, I followed my Plan B and passed the entrance exam to German philology. In addition to German, I also took English and Dutch, and later French and Spanish. I enjoyed learning languages, but it was learning *about* language (i.e. linguistics) that truly became my fascination. After two years of German philology, I found a way to combine linguistics and my old interest in translation and switched to the Applied Linguistics program with a specialization in translation and interpreting. This was the beginning of my professional involvement in translation.

3. Academic translation

Students usually look for opportunities to make money. I used to offer private German lessons and taught German courses, but it was translation jobs that I liked most. I made some money translating materials for the Polish television about Dutch and Belgian TV programs for the hearing impaired. When the texts came from the Netherlands, I was paid for translating from Dutch; when they came from Belgium, I was paid for translating from Flemish, thus adding another language to my set of skills. I also made some money translating academic articles into German. There was a particular Polish philosopher whose papers I translated on a couple of occasions. His style in Polish was antithetical to a famous quote from Wittgenstein: "What can be said at all can be said clearly." (1922: 23) I spent long hours attempting to *interpret* his texts (i.e. trying to understand them so that I could translate them). Years later, I read a review of one of his most difficult-to-follow papers

in my German translation and I remember the reviewer saying that the paper was "written very clearly…" It was then that I knew, that in fact, I did a terrible job as translator. Whether I did a good job as an interpreter remains to be seen, as I am not at all sure I interpreted the text correctly.

4. Translation as a test for the lack of content

As a student of translation, I had also experienced first-hand how completely empty communist political texts were. Before, I simply knew it intuitively. Now, I had experimental proof. Translating texts from the East German communist paper *Neues Deutschland* into Polish went relatively well as my vocabulary in my native language was able to provide a sufficient number of synonyms. However, when translating similar texts from the Polish *Trybuna Ludu* into German, I would get stuck after the second or third reiteration of the same message dressed up in slightly different words; in how many ways could you praise the glorious Red Army liberating Poland from Hitler's tyranny? (cf. Stroińska 2023: 140-141)

5. Literary translation

My proper work as a translator started when I was asked to translate the new edition of Victor Klemperer's 1946 book *Language of the Third Reich* into Polish (Klemperer 1946/1982). It was 1982, the beginning of the martial law introduced by the Polish communist government in December 1981. While working on this translation, commissioned by one of the underground (i.e. illegal) publishing houses, I immediately faced a dilemma: Nazi propaganda in German was in so many ways like the Polish communist propaganda of the 1980s, that translation of some terms came easily. Papers in Poland were full of similar phrases. Everything was *historical*, everything was a *struggle* (i.e. *Kampf um die Ernte – walka o żniwa* – 'a struggle for harvest'). Yet the two political systems, Hitler's Germany and martial law era communist Poland, were very different and one should not try to obscure that distinction by translating Hitler's language into contemporary Polish newspeak. By the time I finished my translation, the underground publisher could no longer publish the book and so I learned about the precarious nature of making money as a literary translator. My Polish translation of Klemperer's book finally came out in 1992 in Toronto (Stroińska 1992).

This, however, was not the end of my adventures with Victor Klemperer. When organizing the Ontario High School German Contest in my department at McMaster University in the late 1980s or early 1990s, I met Professor Manfred Richter, then, the Chair of the Germanic and Slavic Languages at the University of Waterloo. While chatting with him about our research interests, I mentioned Victor Klemperer who turned out to have been the neighbour of Professor Richter's parents in the suburbs of Dresden. What an incredible coincidence!

6. Forming an attachment to the author

Here, I need to explain the effect that translating Victor Klemperer's war-time diary had on me. When you work so closely on someone's text, you may unconsciously start to take on some of their ways of thinking, thus forming an emotional bond with the author. Dealing with translation of accounts of traumatic events is particularly impactful. This is why, in 1994, I travelled to Dresden, then, in the newly reunited Germany, to look for traces of Klemperer's life described in his memoirs. I visited the addresses of the so-called *Judenhäuser* where he and his non-Jewish

wife had to move after Jews were forced out of their apartments and houses. I found the magnificent poplar tree described by Klemperer, the *Babisnauer Pappel* 'the poplar tree of Babisnau,' standing in the middle of the fields. Klemperer reported that local people believed its blooming predicted peace (Klemperer 1982: 71). I found Klemperer's grave in a small cemetery in the suburb of Dölzchen, where he and his wife Eva built a small house before WWII. When the war was over, they were able to return there. I had the address from Professor Richter. The new owners kindly allowed me to come inside and have a look. They also asked me whether I contacted Mrs. Klemperer. I was surprised as I knew Mrs. Klemperer had died in 1951, and I had just seen her grave. Yet it turned out Victor Klemperer did remarry. In 1952, he married his teaching assistant Hedwig, several decades his junior. She was still alive in 1994, and I found her in the Dresden telephone directory. We met and later corresponded until she passed away in 2010. It was Hedwig Klemperer who helped me to understand many aspects of Victor's work, especially the shifting of his political views after the end of the war.

My personal obsession with Klemperer as a person, his life, his thinking, his experiences, was nothing unusual for translators. When I was a graduate student in Edinburgh, I was introduced, in 1984, to a Professor of German, Derek Bowman. Derek translated into English *The diary of Dawid Rubinowicz*, published in 1981. The book is somewhat similar to *Ann Frank's diary*, written by a 12-year-old Jewish boy from a small village in Poland during WWII. He writes about his everyday experiences, fear of the Germans, persecution, and deportations. When he was 15, he was transported to a concentration camp where he perished. The book is written in simple language but the suffering of the young boy is heartbreaking.

In 1984, when we met, Derek Bowman was beginning to suffer from some psychological condition that demonstrated in a false belief that the diary might have been a fraud and that he had been taken advantage of through his personal involvement and emotional attachment to little Dawid Rubinowicz. The story had a tragic ending: one evening, Derek stepped out of the house and onto the nearby train tracks and was killed by the oncoming train. While his obsession was unfounded, I can relate to the feeling of disappointment if I were to believe that Victor Klemperer's diaries were fraudulent. As translators, we often invest an enormous amount of emotions into our work, especially when translating narratives of traumatic experiences. I was dismayed when I discovered that the English translation of the full edition of Klemperer's WWII diaries (Klemperer 1998) was significantly cut. What the English translator, Martin Chalmers, eliminated were mostly Klemperer's observations about Nazi ideology perverting the German language and leading to its degeneration, the very thing that kept Victor Klemperer's spirit alive through the horrors of the war. For him, writing down observations about changes in the German language was his "balancing stick" that allowed him to keep going (Stroińska 2001).

More recently, I became involved in a project of David Hitchcock, a colleague from the Philosophy Department at McMaster, who wanted to undertake a new translation of an important article by Alfred Tarski (1900-1982). Tarski was a Polish logician who worked out a scientific definition of the concept of truth. The existing English translation of Tarski's 1935 Paris conference paper was based on a German translation from Polish (i.e. it was a translation of a translation). David wanted to complete a new translation directly from Polish into English. We worked painstakingly, deliberating every word and every metaphor (the language of science is full of them but we usually overlook them focusing on the message). We were amazed how Tarski, whom we believed to have written the German version himself, was consistent in using the same imagery across the two languages, for example, in Polish: *uprawiać naukę* and in German: *eine Wissenschaft treiben,* both mean 'to cultivate science,' with agricultural associations. We were

now trying to capture the same meanings in English. This was the first time I was working with three languages at the same time, and it was often confusing but also exhilarating, a bit like simultaneously looking at the same object (the message) from three different points of view. After translating one article (Hitchcock & Stroińska 2002), we later undertook the translation of another paper by Tarski, presented at the same 1935 Paris conference. This time, we started to have doubts whether Tarski indeed wrote both versions himself because he declined the invitation to translate into German his longer paper on the concept of truth for publication in the Polish journal *Studia Philosophica* (Tarski 1936) claiming insufficient knowledge of the language. In doing research on Tarski's background, trying to establish who translated into German the paper we were going to translate into English, we found out a lot about Tarski's life and work and I experienced another "Klemperer effect." With the advances in technology, I did not have to travel to Vienna or Lwów in order to walk the streets Tarski walked or see the house where the recipient of his letters, Professor Kazimierz Twardowski, lived. Google Maps allowed me virtual tours through the cities and even made it possible to trace Tarski's hiking trips in the Tirol mountains and in the Pyrenees. Looking up his scientific collaborators, it was heartbreaking to discover that several of them died in the war that was approaching. Some were killed by the Nazis, like Leopold Blaustein, the German translator of Tarski's Magnum Opus on truth. Some committed suicide. This interweaving of life and work that we were now trying to interpret felt very rough. Instead of doing just a new translation of an influential paper whose existing English translation was a translation of the German version, we ended up with a whole book examining the very process of translation of Tarski's philosophical work on scientific semantics (Hitchcock & Stroińska 2025).

7. Court interpretation challenges

After coming to Canada in 1987, I almost immediately got involved in court interpretation. At that time, Ontario courts did not require any kind of certification from the translators. I was recommended by the Polish parish priest in London, Ontario, who was simply too busy to do it himself and so I started going to court, first just in London, then also to other court houses in the area. At some point, the Ministry of the Attorney General introduced certification exams, and I received certification for Polish and German. Apart from earning some money, this activity was also an interesting linguistic exercise. I had clients who were accused of a variety of transgressions, from minor offences up to a few murder cases. Each new person was a new challenge. I quickly learned that when asked to serve as a German interpreter I had to make sure I was going to deal with High German and not Low German spoken by the Mennonites. I once ended up doing my very best to interpret for a Mennonite father who was stopped while driving a car without a permit (Mennonites as a rule do not drive) because the judge insisted that the matter had to go ahead even without a qualified interpreter of Low German. I am glad the person I translated for wasn't convicted.

On one occasion, however, I am sure I contributed to the conviction of a Polish gentleman who was accused of stealing a package of cigarettes from a large store. His line of defence was that he had to put the cigarettes into his pocket as he needed both of his hands to handle the heavy windshield wipers. It took me a while to figure out that when talking about *wycieraczki do samochodu*, literally 'wipers for the car,' he was referring to the rubber floor mats, while I kept talking about windshield wipers, which in Polish would also be *wycieraczki do samochodu*. This gave me an interesting insight into how the Relevance Theory may not always help in deciding which interpretation is the most relevant one. Proposed by Sperber & Wilson (1986), Relevance Theory offers a framework for the interpretation of utterances suggesting that the hearer should be

able to determine which of the competing options is the most relevant one. However, the choice is based on the knowledge of the context of the communication. In many cases of simultaneous or consecutive interpretation, the interpreter does not know the full context.

In my first murder case, the accused, when informed that his wife had died, kept asking in Polish *ale co się stało?* 'but what happened?' The prosecution believed that he must have been the murderer as he never asked *how* his wife had died. I put on my linguist's hat and explained that, in Polish, 'what happened' was a legitimate way to inquire about more details of the situation, and people personally affected by the death would be unlikely to ask 'how did she die?'

Another issue I had to face as a court interpreter was translating vulgar language and obscenities used by clients when addressing the judge or the justice of the peace. As an interpreter, it is my duty to translate, as faithfully as possible, even the most offensive language, but I found it very difficult to do it talking to a person of authority in a full courtroom. In the worst case that I had to deal with, my – probably wrong – way to get out of the unpleasant situation was to add, at the end of each translated turn and with a somewhat different intonation, the proper form of address: *Your Honour* or *Your Worship.* I wanted to make sure that I distanced myself from what I was actually saying. Fortunately, clients who used such offensive language would usually be quickly stopped and removed from the courtroom. This was quite likely the very outcome they were seeking so that their court appearances had to be adjourned time after time.

The COVID-19 pandemic forced court proceedings to be moved online, and this new remote format created additional problems. Court interpreters have been particularly affected by the changes. The inability to communicate face to face and having to deal with significant levels of noise in court proceedings on Zoom may affect the interpreter's comprehension. Generally, if the interpreter cannot hear well what was said, they can ask for a repetition. However, there are situations when we think we comprehend the message because it *makes sense* in the context (seems *relevant*), but our assumed comprehension may turn out to be wrong. In one recent case, the accused was charged with an "attack with a weapon" and the charge was further explained as "hitting the plaintiff with a piece of __ood." Even though I did not hear the onset of the last word, I made what I considered to be the most logical (*relevant*) guess and translated the weapon as a *piece of wood* only to be immediately stopped by the defendant who said the charge was a lie. It turned out he hit someone with a piece of *food*, and specifically a piece of cake, definitely not a prototypical example of a weapon (cf. Kliffer & Stroińska 2005; Stroińska & Drzazga 2018). For a native speaker of English, */fuːd/* and */wʊd/* also differ in the quality of the vowel but this difference may be difficult to observe for a non-native speaker. As remote court proceedings facilitate many legal and logistic problems (like attending multiple court locations in one day and not having to physically bring defendants who are in detention to courthouses), the new format is here to stay and interpretation problems on Zoom (or any other future platform) need to be given more attention.

8. Language brokering

I had the privilege to work with two students – one undergraduate (Maneetpal Badesha) and one graduate (Joyce Zheng) – on projects involving language brokering. This is the practice, widespread in multilingual and multicultural countries, that attracts immigrants, when immigrant children who usually learn the new language faster, are asked to interpret for their parents and grandparents in official situations. In Translation Studies, the term *community interpreting* is used (cf. Hale 2007) for a similar phenomenon but community interpreters are normally adults who help

in situations where someone who does not know the official language needs to communicate with health care providers, government services, etc., while language brokering involves young children and teenagers with no training, no professional preparation, who usually are not mature enough to assist in such matters. While I have never experienced language brokering myself, I appreciate the heavy burden of responsibility that is related to the task of translating in contexts that are usually way beyond the child's maturity and language competence. In Canada, using children to interpret for their parents and other family or community members is discouraged; however, it still happens, for example, in medical and other emergency situations. My students researched the effects of childhood language brokering on family relationships, overall linguistic competence in both heritage and host country's languages, and future career choices. A lot more research needs to be done on the practice and effects of language brokering. One finding that really amazed me was how much the very fact that someone wanted to do research on their childhood experience meant for the young people who did broker for their families. They were surprised that there was *a name* for what they did, that it was a widespread phenomenon, and that linguists were interested in how it affected them. Most language brokers also said that even if they resented the task as children, they would still do it if they have a choice. It became part of their identity.

It could be an interesting task to compare language brokering and community interpreting from the perspective of both applied linguistics and translation studies.

9. Teaching translation: theory and practice

Over the years, I have taught translation classes in our German language program and later introduced a Translation Theory course into our linguistics and cognitive science language programs. When I was a student myself, Translation Theory was one of my favourite courses. In particular, the idea of linguistic relativity – formulated by Sapir and Whorf (but originating in the German language philosophy of Herder and Humboldt) – helped me to better understand the effects of political ideology on language and then, in turn, the effects of such manipulated language on the perception of reality. While I found the work of Sapir and Whorf illuminating, I never believed that we truly cannot communicate across the barriers of natural languages. However, political ideology and propaganda based on language distortion were able to shape the perception of reality. This is, unfortunately, still the case, maybe even more today than it was in the 1970s when I was an undergraduate student (cf. Stroińska 2023).

The *teaching translation* practice has always been fun for me, and I sincerely hope that I was able to pass on some of my enthusiasm to my students. Translating from German often feels like solving a puzzle – the morphology of the language usually gives clues as to which noun phrase is the subject and which is the object. English speakers tend to assume that the first noun phrase in the sentence must be the subject but German offers more flexibility and so the direct object may be put into the sentence initial position if it is the focus of attention. The articles and adjectives are marked with case endings that point to their syntactic roles and help disambiguate complex structures. I have learned that the best texts for translation exercises were those containing jokes as understanding the punch line served as a reward for good translation. Students knew that if there was nothing funny about the story they translated, they must have missed something (cf. Stroińska & Drzazga 2015). I also used examples of *translating* literary texts into films, using both good and not so good examples (Stroińska 2007). We would, for example, read a chapter from Lewis Carroll's *Alice in Wonderland* in the 1865 English original, then its German version with masterfully recreated humour and puns in the excellent 1869 translation by Antonie Zimmermann.

For example, when Alice meets the Mock Turtle and Gryphon, the Mock Turtle tells her about his school timetable, where they had fewer lessons on each subsequent day of the week. When Alice seems surprised with such a novel setup, the Mock Turtle explains: this is why they are called *lessons*, because they *lessen* from day to day. This pun would not work in German. Antonie Zimmerman came up with an original solution: the number of students in the class decreased day by day. The explanation was: this is why schools employ teachers (*Lehrer*) because they empty (*leeren*) the classroom. Both words share the same pronunciation with a long *[e:]* and the pun is equally funny in German as the original pun was in English. Lewis Carroll was a logician at the University of Oxford and his bizarre humour was achieved by bending literal meanings of words and exploiting logical reasoning ad absurdum. Once the students understand and appreciate the original humour in a fragment of *Alice in Wonderland*, for example, in the chapter about the "mad tea party," we would watch a corresponding fragment of Disney's 1951 animated movie classic. The film is addressed to children while Lewis Carroll's book only looks like a children's story. The students are usually astonished by the fact that very little of Carroll's sophisticated wit actually made it into the movie, making it a bad example of *adaptation* from one genre (a literary text) into another (a film). Absurdist humour is not the same as silliness.

10. Cognitive aspects of translation

In every culture, there are some keywords (i.e. concepts) that are more prominent than others. Growing up in post-war Poland, I was virtually surrounded by references to *death*. In Warsaw, the city that was almost completely (over 90%) destroyed, nearly every street has plaques reminding passers-by that they are walking past "places of execution sanctified with the blood" of innocent civilians. As a child, I spent considerable time visiting cemeteries and learned about my country's history by recognizing the significance of certain dates; young men who died in 1920, must have been killed in the Polish-Soviet war; dates in the 1939-1945 bracket were the victims of the WWII; 1944 deaths were likely the result of the Warsaw Uprising; war dates on symbolic graves suggested death in German concentration camps or in Soviet deportation. Some of these events, for example, the war of 1920 or Soviet deportations during WWII, were taboos in communist times and were not taught about in school.

The images of death were common in literature and in visual arts. Death in Polish was often personified as a woman, either young and beautiful or old but caring rather than cruel. It was not unusual to read about Death lovingly embracing the dying. When I started to learn other languages, I realized that Germanic Death was visualized rather differently. The Grim Reaper, a skeleton in a tattered garb, was a very different concept than the Slavic Death. Both Slavic and Romance languages use feminine gender to refer to death and this grammatical feature affects the conceptualization. The same is the case for the noun *war* (Polish feminine *wojna*), also imagined as a beautiful woman who is tempting young men to blindly follow her, as described by the lyrics of a popular song.

My doctoral student and now colleague and collaborator, Grażyna Drzazga, researched the concept of grammatical gender in translation, focusing on its cultural, grammatical, and cognitive aspects (cf. Drzazga & Stroińska 2012; Drzazga, Stroińska & Sullivan 2012; Stroińska, Drzazga & Kurowska 2014). Polish gender agreement patterns are complex but quite mechanical: the verb needs to agree with the grammatical gender of the subject. Thus, in Polish, Cinderella (masculine *Kopciuszek*) takes masculine forms of the verb and of the adjectives. The recent debate on gender neutral language for non-binary individuals takes a special meaning in Polish where many parts of

speech require gender agreement. However, in Polish, gender in language is more a grammatical rather than a social construct.

11. The language of Populist politics

As a student of translation, I struggled with the empty political discourse of the communist era. Now, more than 40 years later, I am in awe about the new quality of populist political discourse and, especially, by language used by the former US president and 2024 presidential candidate, Donald Trump. Incidentally, the first to notice and comment on this striking phenomenon of Trump's language were the translators and interpreters of his speeches. Already in December 2016, Claire Levenson (2020) pointed out that "[w]ith his incoherent, repetitive style, coarse expressions, and dubious syntax, the American president has made life difficult for French translators and journalists." She called translating Trump a "mission impossible." There are many particular reasons for the difficulty in translating Donald Trump. The first has to do with the syntactically incomplete, interrupted sentences and sentence fragments that may escape the critical attention of the local audiences but are impossible to translate in any meaningful way. Let's take, as a random example, the following fragment of Trump's 2020 Rose Garden speech:

> We have a very strong travel ban and we don't want people that are going to come in and blow up our cities, do things, and frankly with the – with the liberal Democrats running the cities that we do have – where they do have problems, maybe they wouldn't mind but I would mind and the people of this country mind. (Cillizza 2020)

As pointed out before, the interpreter needs to understand the message in order to translate it into another language. In machine translation there is a saying: "garbage in, garbage out." The Oxford English Dictionary defines it as "incorrect or poor quality input will always produce faulty output" and attributes it to the mid-20th century computing studies. In translating texts like the fragment above, a faithful translation would result in garbled nonsense (not unlike the original). Some translators might try to translate something that is hardly comprehensible or simply non-standard into a more coherent message. My dear friend, the late Dr. Zofia Klimaszewska, the founder of Dutch Studies at Warsaw University, travelled as an interpreter with President Lech Wałęsa to The Netherlands. Wałęsa was a technician in the Gdańsk shipyard, the leader of the Solidarność movement and a political prisoner during the martial law period in Poland. He became the president of Poland after the fall of the communist system and was famous for his colourful language and interesting expressions. His message was usually clear and to the point but often expressed in non-standard ways, as in, for example, his famous statement *jestem za a nawet przeciw* 'I am for and even against.' People who listened to Wałęsa in Polish sometimes did not like his simple language. Those who listened to him in translation heard a sanitized version, without any linguistically awkward phrases.

However, translating non-standard language into standard language is very different from trying to make a coherent message from something that makes little sense. This is why some interpreters translate Donald Trump's speeches word for word and add disclaimers, a bit like my adding *Your Honour* to the offensive language of my courtroom clients. According to Levenson (2020), Bérengère Viennot is an example of a French interpreter who "dares to translate Trump." This means accurately relaying dubiously constructed phrases, such as when the president mixed up his words ("mentality" instead of "immunity") while referring to herd immunity to COVID-19: "Over

a period of time, sure, with time, it goes away and you'll develop, you'll develop like a herd mentality — it's going to be herd-developed, and that's going to happen. That will all happen." Viennot (2019) touches also on the other problem with Trump's language: the poverty of his vocabulary and repetition of certain words (like *great* or *excellent*). Instead of making it easier, it makes translation more difficult as some common words in English (i.e. *great*) may have to be translated differently in different contexts. Viennot also faithfully translated into French expressions such as "shithole countries" (*pays de merde* or *trous à merde*) in Trump's angry anti-immigrant outbursts. The difficulties with translating Donald Trump's language point to problems with his message. To repeat Ludwig Wittgenstein's statement in his *Tractatus Logico-Philosophicus:* "What can be said at all can be said clearly; and whereof one cannot speak thereof one must be silent." (1922: 23) While we may not easily notice the emptiness of Trump's discourse when we hear it, it is impossible not to stumble over it when we attempt to translate it. Thus, the process of translation can serve as a useful detector for phony language and convoluted messages.

12. Conclusions

Over the four decades of various encounters with translation theory and practice, I have learned a number of valuable lessons. Let me finish this paper with a short summary of these insights.

- From my experience with translating communist newspaper texts as a student of translation, I have learned that it is difficult to say the same thing in different ways especially when the content is meagre.
- From my work on translating Victor Klemperer's book on the language of the Third Reich, I have learned that when politicians from different sides of a conflict are saying the same thing, it may still mean different things.
- From my work on translating Victor Klemperer and Alfred Tarski, I have learned that translators may get incredibly attached to the authors they are translating.
- From supervising students who researched children as language brokers, I have learned how involving children in translation may influence their linguistic development in both languages and shape their family's dynamic.
- From working on translating grammatical features such as gender, I have learned that they may shape the perception of cultural concepts such as *death* or *war*.
- From all these experiences, I have learned that every translation is an act of interpretation. You cannot translate what you cannot understand. As such, the act of translating also serves as a perfect *bullshit detector*. Translators and interpreters are like canaries in a coal mine. They may be the first to detect that something is seriously wrong with the language of a political leader or a political formation.

Translation is an incredible, multi-faceted activity of the human brain and mind that deserves more attention in linguistics and in the cognitive study of language.

References

Badesha, Maneetpal. 2022. *The long-term effects of language brokering*. Hamilton: McMaster University. (Undergraduate Thesis)

Bowman, Derek (trans.) 1981. *The diary of David Rubinowicz*. Edinburgh: William Blackwood.
Carroll, Lewis. 1865. *Alice's Adventures in Wonderland*. London: Macmillan.
Carroll, Lewis. 1869. *Alice's Abenteuer im Wund*erland. Translated by Antonie Zimmermann. London: Macmillan and Comp.
Cillizza, Chris. 2020. *The 45 most incoherent lines from Donald Trump's rambling Rose Garden speech. The Point*. CNN Politics online. (https://www.cnn.com/2020/07/15/politics/president-donald-trump-rose-garden-speech/index.html#:~:text=29)%20)
Drzazga, Grażyna & Stroińska, Magda. 2012. The grammatical gender of Death: A textual and discourse approach. *Text und Diskurs* 5. 324-340.
Drzazga, Grażyna & Stroińska, Magda & Sullivan, William J. 2012. Pratchett's death in translation: Masculine śmierć? In Wawrzyczek, Irmina & Kędzierska, Aleksandra (eds.), *(Non)omnis moriar: Cultural and Literary Discourses of Death and Immortality*, 83-92. Lublin: UMCS Press.
Hale, Sandra B. 2007. *Community Interpreting*. New York: Palgrave Macmillan.
Hitchcock, David & Stroińska, Magda. 2024. *Alfred Tarski on scientific semantics* (Logic, Epistemology and the Unity of Science, vol. 62). Cham: Springer Nature.
Klemperer, Victor. 1946. *LTI: Notizbuch eines Philologen* [Language of the Third Reich]. Leipzig: Verlag Philipp Reclam jun.
Klemperer, Victor. 1982. *LTI: Notizbuch eines Philologen* [Language of the Third Reich] 7th edn. Leipzig: Verlag Philipp Reclam jun.
Klemperer, Victor. 1995. *Ich will Zeugnis ablegen bis zum letzten: Tagebücher 1933-41.* Nowojski, Walter (ed.), in cooperation with Klemperer, Hadwig. Berlin: Aufbau Verlag.
Klemperer, Victor. 1998. *I shall bear witness: The diaries of Victor Klemperer, 1933–41*, translated by Chalmers, Martin. London: Weidenfeld & Nicolson.
Kliffer, Micheal & Stroińska, Magda. 2005. Relevance Theory and translation. *Linguistica Atlantica* 25. *165-172.*
Levenson, Claire. 2020. Translating Trump: Mission Impossible. *France-Amérique*. November 4, 2020. (https://france-amerique.com/translating-trump-mission-impossible/) (Accessed November 4, 2020).
Sperber, Dan & Wilson, Deirdre. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.
Stroińska, Magda. 1992. *LTI: Notatnik filologa*. Toronto: Polish Publishing Fund in Canada. (Translation of Klemperer 1946)
Stroińska, Magda. 2001. Beyond language and culture: relative points of view (preface and introduction). In Stroińska, Magda (ed.), *Relative points of view: linguistic representation of culture*, ix-xii and 1-16. Oxford: Berghahn Books.
Stroińska, Magda. 2001. Review of V. Klemperer's book 'I will bear witness. A Diary of the Nazi Years 1933-1941.' *Seminar* 37(3). 271-273.
Stroińska, Magda. 2007. The adventure of teaching translation as intercultural communication. In Rieger, Caroline L. & Plews, John L. & Lorey, Christopher (eds.), *Interkulturelle Kompentenzen im Fremdsprachenumnterricht. Intercultural Literacies and German in the Classroom: Festschrift für Manfred Prokop,* 51-70. Tübingen: Gunter Narr Verlag.
Stroińska, Magda. 2023. *My life in propaganda: Language and the totalitarian regimes*. Calgary: Durvile and UpRoute.
Stroińska, Magda & Drzazga, Grażyna. 2015. Punchline as reward: How to use humour in teaching translation. In Löschmann, Martin (ed.), *Humor im Fremdsprachenunterricht* (Deutsch als

Fremdsprache in der Diskussion vol. 10), 261-276. Frankfurt am Main: Peter Lang.
Stroińska, Magda & Drzazga, Grażyna. 2018. Relevance Theory: Interpreting and translation. In Malmkjær, Kirsten (ed.), *The Routledge Handbook of Translation Studies and Linguistics,* 95-106. New York: Routledge.
Stroińska, Magda & Drzazga, Grażyna & Kurowska, Katarzyna. 2014. Translating grammatical gender: Current challenges. *Studia o Przekładzie* [Translation Studies] 36. 175-190.
Stroińska, Magda & Hitchcock, David. 2002. On the concept of following logically (new annotated English translation of Alfred Tarski's 1936 paper with introduction). *History and Philosophy of Logic* 23. 155-196.
Viennot, Bérengère. 2019. *La Langue de Trump*. Paris : Les Arènes.
Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*, Ogden, Charles Kay (ed.). Translated by Frank Plumpton Ramsey. London: Kegan Paul, Trench, Trubner & Co., Ltd. Republished as ebook in 2010 as part of the Gutenberg project. Available at: https://www.gutenberg.org/files/5740/5740-pdf.pdf
Zeng, Joyce. 2024. Two birds, one stone – idioms across languages and the language brokering effect. (Paper presented at the 50th LACUS conference at the University of Waterloo/St. Jerome's University, Waterloo, July 2024)

This article was first published at https://www.lacussquare.org

LACUS Forum **50:1.13-26 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

The relationship between articulation rate and utterance length in varieties of Canadian French

Wladyslaw Cichocki[1], Svetlana Kaminskaïa[2] & Luke Hagar[3]
[1]*University of New Brunswick*, [2]*University of Waterloo*, [3]*McGill University*

Abstract: This paper explores the association between articulation rate (AR) and utterance length – defined as the length of inter-pause intervals (IPIs) – in regional varieties of Canadian French. Data are from readings of a text by speakers of *québécois* (Québec City), *acadien* (Tracadie, New Brunswick), and *ontarien* (Windsor, Ontario) varieties. Results show that, in addition to region and speaker age, IPI length is a significant predictor of AR. However, there are no significant interactions between IPI length and region or the other external predictors. Additional analyses indicate that regional differences in IPI length are significant, but these do not appear to underlie differences in AR. A follow-up exploratory study of mean vowel duration reveals that, at slower ARs, correlations between mean vowel duration and IPI length are different for Québec City speakers when compared with speakers from the other regions; at faster ARs, these correlations are nearly identical across the three regions.

Keywords: articulation rate, Canadian French, dialect variation, inter-pause intervals, mean vowel durations

Languages: French

1. Introduction

ARTICULATION RATE (AR) IS A MEASURE OF SPEECH TEMPO – how quickly or slowly a speaker is talking. It is measured as the number of units uttered in a specific timeframe, excluding pauses and disfluencies. AR is usually expressed as the number of words per minute, phones per second or syllables per second; an equivalent measure is average syllable duration (in msec/syll). A separate but related temporal measure known as 'speech rate' is used to analyze speech tempo that includes pauses.

Different factors – including region, speaker age, gender, and style – can influence variation in AR. Regional variation has been described for English in the United States (Jacewicz et al. 2010; Kendall 2013; Clopper & Smiljanic 2015), for French in Belgium, Canada, France, and Switzerland (Avanzi et al. 2012b; Kaminskaïa et al. 2016), for Dutch in the Netherlands and Belgium (Verhoeven et al. 2004; Quené 2008), for Spanish in Argentina and Spain (García-Amaya et al. 2023), among others. Studies of speaker age have noted that younger speakers tend to articulate faster than older speakers (Bóna 2014). However, results for gender do not follow a consistent pattern: some studies show that males have faster ARs (Verhoeven et al. 2004; Kendall

2013), others show that this effect depends on style (Jacewicz et al. 2010; Schwab & Avanzi 2015), yet others find no effect of gender (Clopper & Smiljanic 2015). Similarly, style can show opposite effects: some studies report faster ARs in reading than in spontaneous speech (Simon et al. 2010), while other studies report that ARs are faster in spontaneous speech (Schwab & Avanzi 2015).

1.1 About utterance length

The main internal factor that appears in studies of AR variation – and this is the focus of the present paper – is often referred to as "utterance length." The relevance of this factor was noted in two early studies about French. Malécot et al. (1972) defined an utterance as the stretch of speech, two syllables or longer, between two pauses. The authors measured ARs across these stretches to study recordings of the conversations of 50 speakers of upper-middle-class Parisian French. One of their main observations was the existence of a positive correlation between AR and utterance length: "the longer the utterance, the faster the AR, and vice versa" (p. 235). Grosjean & Deschamps (1975) used a similar definition of utterance, referring to it as a "run" or "suite sonore" (p. 150). Calculating ARs across these small "runs" provided the researchers with descriptive statistical information about the mean, standard deviation, and distribution of ARs. Comparison of English and French radio interviews showed both similarities and differences. On the one hand, English and French had similar ARs (medians are 5.0 syll/sec vs. 5.21 syll/sec) with very similar distributions of AR values; on the other hand, English had shorter utterance lengths than French (9.5 syll/unit vs. 12.0 syll/unit) (p. 160), and English also showed a stronger correlation between AR and utterance length than French (0.35 vs. 0.23) (p. 156).

These analyses made several important contributions to our knowledge of AR. First, they revealed the existence of 'within-speaker' AR variation and showed that this local variation is very large. Until that point in time, researchers had worked with global or 'between-speaker' measures of AR – that is, they calculated ARs over each speaker's entire interview or reading passage – and thus they ignored or missed local, 'within-speaker' variation. Second, there was the recognition that 'between-speaker' AR variation was larger than had been previously assumed (cf. the work on pauses by Goldman-Eisler 1968). Third, the studies brought attention to the relationship between AR and utterance length; notably, Grosjean & Deschamps (1975: 160) suggested that utterance length may be a relevant factor in explaining AR differences between English and French.

More recently, developments in statistical modeling have permitted researchers to include both between-speaker and within-speaker factors in the analysis of AR variation. Quené (2008) used multi-level modeling to analyze the spontaneous conversations of 160 speakers of Dutch from the Netherlands and Belgium. He found that two factors have a significant effect on AR: a between-speaker factor, region – speakers from the Netherlands have faster ARs than those from Belgium – and a within-speaker factor, utterance length (which he called "phrase length"), noting that "speech tempo increases with phrase length" (p. 1109). The same pattern had been observed by Malécot et al. (1972), as noted above, and this observation has been replicated in studies of American English (Kendall 2013; Clopper & Smiljanic 2015).

The multi-level statistical modeling also pointed to another possible explanatory role of utterance length. Quené (2008: 1111) suggested that between-speaker effects on AR of factors such as region and age might be attributed to between-speaker differences in phrase length. For example, older speakers produce shorter phrases than younger speakers, which could explain why ARs become slower as speaker age increases. In addition, at the regional level the pattern of the relation between phrase length and AR is reversed: speakers from the Netherlands produce shorter

phrases with significantly faster ARs than speakers from Belgium who have longer phrases but slower ARs. The hypothesis that phrase length underlies or explains regional and other differences observed in AR variation received further support in a study of regional varieties of American English by Jacewicz et al. (2010). However, Schwab & Avanzi (2015) found no support for this hypothesis in their study of European French.

An additional aspect of the effect of utterance length on AR is the possibility that utterance length can interact with other predictors. In their study of reading and spontaneous speech in seven varieties of European French, Schwab & Avanzi (2015) uncovered significant interactions of utterance length with region and style; the effect on AR is stronger in some regions and in certain styles than in other regions and styles. This suggests that as speakers modulate their ARs, with higher or lower rates as utterance length varies, the extent of this modulation may be determined by an external factor such as region. However, we are not aware of other studies that find a similar interaction between utterance length and one or more social factors. For example, in a recent study of spontaneous speech in English spoken in Canada, the United Kingdom, and the United States, Tanner et al. (2024) observed that the utterance length effect on AR varies little across 13 speech corpora and 1092 speakers.

1.2 Scope of this study

The present study examines AR in French spoken in three regions of Canada: Québec City (in the province of Québec) where the *québécois* variety is spoken, Windsor (in the province of Ontario) where the *ontarien* variety is spoken, and Tracadie (in the province of New Brunswick) with the *acadien* variety. These varieties represent the two main groups of dialects of French spoken in Canada: the *québécois* and *ontarien* varieties belong to the Laurentian French group, and the *acadien* variety from New Brunswick belongs to the Acadian French group. In terms of sociolinguistic setting, the regions are very different. Québec City is a large urban centre where French is the official language and where most residents (96.6% according to 2001 census figures) are native speakers of French. Windsor is a mid-sized urban centre where Francophones are a small minority (3.9%) who live in a largely English setting and where English is the official language. Tracadie is a small rural town where most residents (96.9%) are native Francophones. Both English and French are official languages in the region, but French is in a minority setting.

In our initial study of AR variation in these three varieties (Cichocki et al. 2024), we analyzed both reading and conversational speech styles. We defined utterance length as the length of an inter-pause interval (IPI length). Results showed that AR varies by region, speaker's age, style, and IPI length. We explained the regional differences in AR – not in terms of dialect differences – but in terms of the degree of English-French language contact: ARs are faster in Québec City, where there is a low degree of language contact, and slower in Windsor and Tracadie, where there are greater degrees of language contact. Other results were consistent with past findings in the literature: AR is faster in spontaneous speech than in reading; AR decreases with age in the reading style but not in spontaneous style; AR becomes faster as the length of IPIs increases.

The present study follows up on this research with a closer analysis of the association between IPI length and AR and with attention to some of the issues noted in previous research. The study is limited to reading style, which ensures that comparisons among speakers are based on the same linguistic content. In addition to reproducing some of the original study's results, we are asking the following research questions:

a) What is the nature of the effect of IPI length on AR? Is the pattern observed in earlier studies – as IPI gets longer, AR gets faster – also present in the three regional varieties of Canadian French? Does IPI length interact with region, as observed in European French varieties?
b) Does IPI length underlie or explain region, age or gender differences in AR? Specifically, can between-speaker effects on AR of region, age or gender be attributed to between-speaker differences in IPI length?
c) What can variation in vowel duration reveal about the relationship between AR and IPI length? In a supplementary study, we explore the relationship between AR and IPI length by examining the distribution of mean vowel durations across the three regions. Part of the motivation of this supplementary study lies in the observation that vowel durations are inherently correlated with syllable durations, and they are thus sensitive to variation in AR.

2. Method

2.1 Speakers and speech materials

Participants were 31 native speakers of three varieties of French spoken in Canada: *québécois*, spoken in Québec City; *acadien*, from Tracadie; and *ontarien,* spoken in Windsor. All participants were born and raised in the region, and they resided there at the time of the recordings. Specifically, 8 were from Québec City, 11 from Tracadie, and 12 from Windsor. There were 14 males and 17 females; their ages ranged from 17 to 86 years. The sample is not perfectly balanced with respect to gender and age across the three regions; there are only two male participants from Québec City. However, the sample has eight participants from the ages of 17 to 60 in each region, indicating that one can use statistical modeling to estimate effects of both region and age on AR.

Speech samples are from interviews conducted in the context of the *Phonologie du français contemporain* (PFC) project, which documents regional and stylistic variation in varieties of French spoken around the world (Durand et al. 2009). The interviews were designed to elicit different speech styles, including a reading task that corresponds to a formal style. Each participant read the text of a made-up newspaper article (the *Le Premier Ministre ira-t-il à Beaulieu?* passage) at a natural speed. The interviews were usually conducted at the participant's home.

2.2 Measurements and statistical analysis

The recordings were analyzed acoustically with *Praat* software (Boersma & Weenink 2022). Segmentation and transcription of phones followed standard visual and auditory criteria (Peterson & Lehiste 1960). Syllable boundaries were established manually and were based on actual pronunciations; that is, attention was paid to speech features such as deletions and insertions of vowels and consonants, vowel fusions, and the presence and absence of liaison consonants. An IPI was defined as the stretch of speech between two pauses, where pauses were considered to be silent intervals of 80 msec or more. This approach allowed us, for example, to assign 80 msec of silence to word-initial plosives, with the rest of the silence being attributed to the pause; for a similar approach, see Šturm & Volín (2023). We note that there is no generally accepted threshold for defining pauses, with researchers using durations that vary from 60 msec to 250 msec (Kendall 2009). In a few cases, filled pauses that directly preceded or followed speech associated with the text served as the boundary of an IPI. Filled pauses, hesitations, false starts, and speech overlaps were excluded from analysis.

Figure 1 provides an example transcription of the excerpt "*et plus récemment, son usine de pâtes italiennes*" from the *Le premier ministre ira-t-il à Beaulieu?* text reading produced by a 25-year-old female speaker of the *acadien* variety. The two uppermost sections are acoustic outputs from *Praat*: the waveform and wideband spectrogram of the excerpt, including the fundamental frequency (in blue, or darker shade) and intensity (in yellow, or lighter shade) contours. The three tiers below these outputs show the segmentation of the acoustic signal into phones, syllables, and IPIs. Note the presence of three silent pauses (marked as "_") which delimit the excerpt into two IPIs: "*et plus récemment*" and "*son usine de pâtes italiennes."* The two lowest tiers document the articulation rate and the IPI duration and length associated with each IPI.

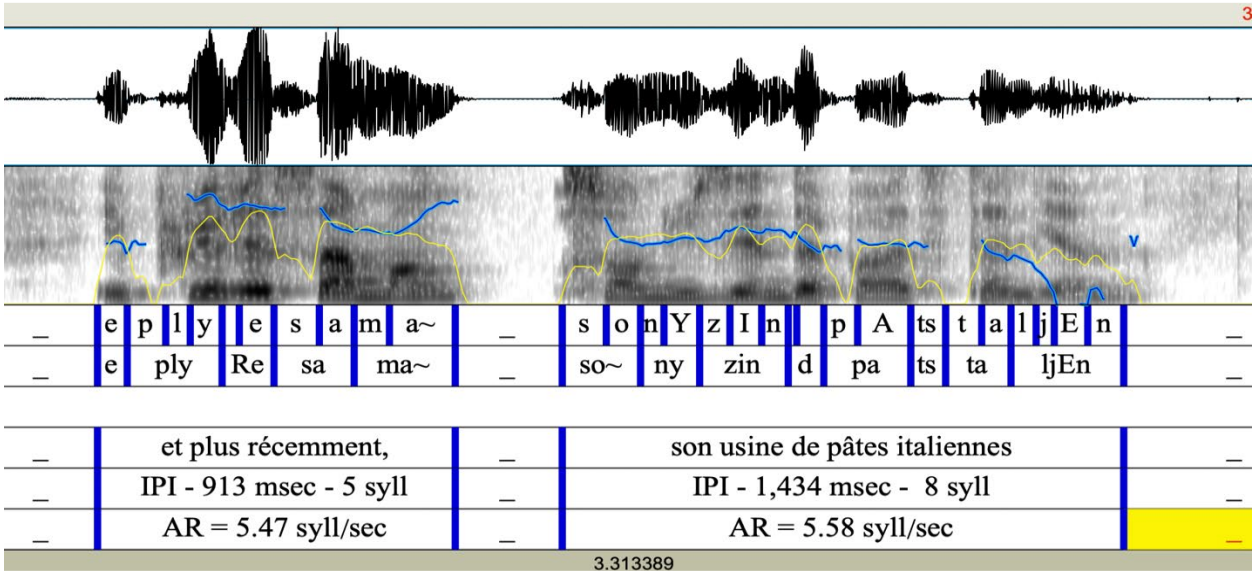

3
_ e p l y e s a m a~ _ s o n Y z I n p A ts t a l j E n _
_ e ply Re sa ma~ _ so~ ny zin d pa ts ta ljEn _
_ et plus récemment, _ son usine de pâtes italiennes _
_ IPI - 913 msec - 5 syll _ IPI - 1,434 msec - 8 syll _
_ AR = 5.47 syll/sec _ AR = 5.58 syll/sec _
3.313389

Figure 1: Example transcription of an excerpt from the *Le premier ministre ira-t-il à Beaulieu?* text produced by a female speaker of Acadian French from Tracadie

Articulation rate, measured as the number of syllables per second, was calculated separately for each IPI. IPI length was calculated as the number of syllables in the IPI stretch of speech. Durations of vowels and syllables, and IPI lengths were measured automatically with *Praat* scripts. The corpus of the text readings by the 31 speakers contains approximately 35 minutes of fluid, connected speech (excluding pauses and disfluencies), with 1318 IPIs and 9808 syllables.

Statistical analyses of these data used linear mixed-effects modeling and ANOVA. The follow-up exploratory study of mean vowel durations was also carried out with visual inspection of various graphics. All linear mixed-effects analyses were made using the *nlme* package in R (Pinheiro et al. 2017; R Core Team 2020).

3. Results and Discussion

3.1. Effect of IPI length on AR variation

To study the first research question about the nature of the effect of IPI length on AR variation, we mirror several parts of our original study of AR variation in these three regional varieties (Cichocki et al. 2024).

Figure 2 gives preliminary descriptive statistics for AR and IPI length. The boxplots in the left

panel of Figure 2 present the distribution of AR by region. They show that average ARs differ across regions: Québec City has a faster average AR than Tracadie and Windsor. The plots also indicate some skewness in the distribution. The combined sample skewness of 0.60 is considered moderate, which often requires a square root transformation. The boxplots in the right panel summarize the distribution of IPI values for the three regions. They show that IPI lengths also differ among the three regions. The combined sample skewness of 2.12 is considered high. To meet the model's symmetry and linearity assumptions, we applied a logarithmic transformation to both AR and IPI.

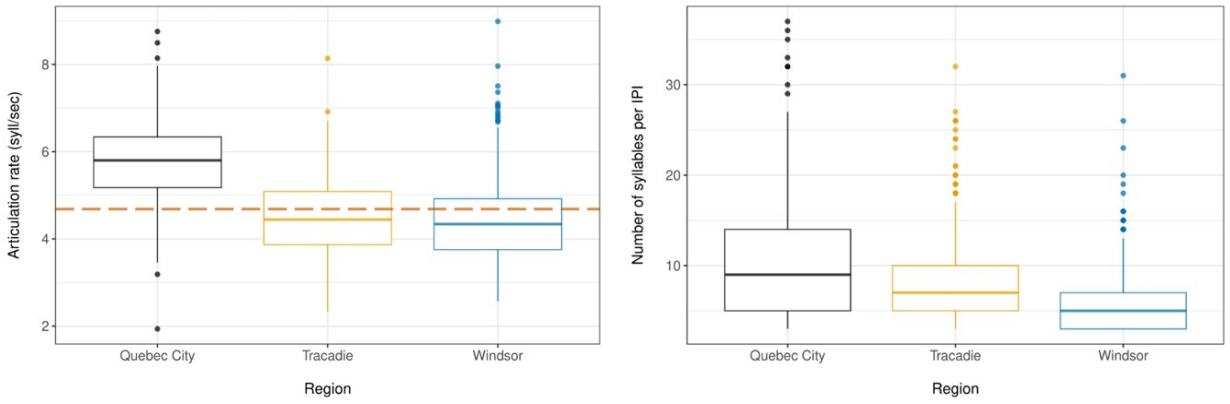


Figure 2: Boxplots for articulation rates (left panel) and IPI values (in number of syllables) (right panel) by region; the dashed horizontal line marks the overall mean

All potential predictors of AR variation – region, age, gender, and IPI length – were included in a linear mixed-effects modeling. These predictor factors and their interactions were treated as fixed effects in the modeling; speakers were a random effect. Logarithmic transformations of AR and IPI lengths (as justified above) were used. Age and the logarithm of IPI were centered in the model-fitting process. The best-fitted model that accounts for AR variation is given in Table 1. The Q-Q plot of the residuals was considered to provide validation for this model.

Table 1: Estimated parameters (with standard error) of linear mixed-effect model of log(AR)

Effect	Estimate	Std Error	*p*-value
Fixed effects			
Intercept (reference: Québec)	1.676	0.035	<.001
Tracadie vs. Québec	-0.185	0.045	<.001
Windsor vs. Québec	-0.189	0.045	<.001
Age	-0.003	0.001	.002
Logarithm of IPI	0.069	0.009	<.001
Random effects			
Intercept	–	0.088	–

Results of the statistical modeling show that IPI length, age, and region are significant predictors of AR. The model reveals several patterns:

1) for region: mean ARs in Québec City are significantly higher than those in Tracadie and Windsor,
2) for age: ARs decrease as speakers' ages increase,
3) for IPI length: shorter IPIs are pronounced with slower ARs, longer IPIs with faster ARs'.
4) gender is not a significant predictor (p = .70),
5) there are no interactions among the predictor factors: region-by-age (p = .94), gender-by-age (p = .89), IPI-by-region (p = 0.65), IPI-by-age (p = .07).

According to this modeling, IPI length has a positive correlation with AR – or, more specifically, there is a direct, linear relation between the logarithmic transforms of IPI and AR values. A graphic representation of this relation, given in Figure 3, shows how log(IPI) and region impact AR. The pattern is a familiar one: shorter IPIs are pronounced with slower ARs, and longer IPIs with faster ARs. The same pattern has been noted in the literature (Malécot et al. 1972; Quené 2008; Kendall 2013; Tanner et al. 2024; among others), and the Canadian French data confirm it. A reasonable explanation that has been proposed by, for example, Quené (2008: 1111) is "anticipatory shortening": speakers shorten their syllables – that is, they increase their AR – if they anticipate a longer IPI.

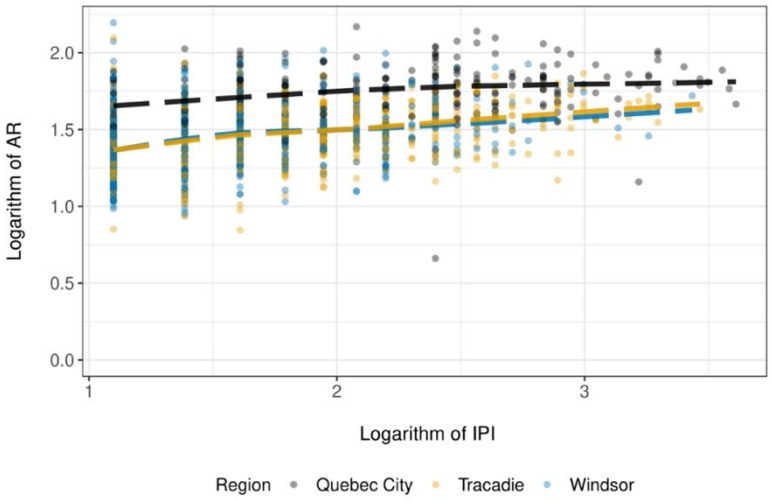


Figure 3: Plot of log(IPI) vs. log(AR), grouped by region

In addition, the statistical modeling indicates that there are no interactions of IPI length with region, speaker age, or gender. Indeed, there are no interactions between any factors in the model. Figure 3 shows that the lines that plot the relation between IPI length and AR for the three regions appear to be very similar and that they are roughly parallel. This graphic evidence points to the absence of a region-by-IPI length interaction, and it lends support to the conclusion that the effect of IPI length on AR is the same across regions.

This conclusion contrasts with the findings reported by Schwab & Avanzi (2015: 101-102) in their study of seven regional varieties of European French. These authors observed interactions of IPI length with region and style. For example, in reading style, the slope of the IPI vs. AR line from their fitted linear model is steeper for some regions (such as Genève and Neuchâtel in Switzerland) than for others (such as Lyon in France and Liège in Belgium). This is an important contrast with the three Canadian varieties, where there is no such regional effect.

3.2 Does IPI length explain the effects of external factors?

Our second research question asks whether IPI length underlies or explains the effects on AR variation of the external factors. Recall that research by Quené (2008) and Jacewicz et al. (2010) argued that utterance length (that is, IPI length) plays a role in explaining regional and age differences in AR. They showed that, in Dutch and US English, between-speaker effects on AR – such as region and age – can be attributed to between-speaker differences in utterance length. To address this question, we apply a three-way ANOVA to the IPI values, with region, age, and gender as independent variables.

Results show that mean IPI length differs significantly by region ($F(2,26)$: $p < .001$), even after accounting for gender and a binary dichotomization of the age variate at the median age of 41 years. Gender ($F(1,26)$: $p = .04$) was significant in this ANOVA analysis – female speakers have longer IPIs than male speakers – but age ($F(1, 26)$: $p = .26$) was not significant. Post-hoc two-way ANOVAs revealed that mean IPI length differs significantly in all three regions after applying a Bonferroni correction: Windsor vs. Québec City ($F(1, 17)$: $p < .001$), Windsor vs. Tracadie ($F(1,20)$: $p = .002$), and Québec City vs. Tracadie ($F(1,16)$: $p = .01$). To satisfy the normality assumption, the logarithm of IPI length was used as the dependent variable in the initial and post-hoc ANOVAs.

These statistical tests provide relevant information for our research question. The regional averages in IPI length (see Figure 2) follow the order: Windsor (5.75 syll) < Tracadie (7.81 syll) < Québec (11.24 syll), with significant differences among the regions. This order parallels – partially – the order of average regional ARs (see Figure 2): Windsor (4.42 syll/sec), Tracadie (4.49 syll/sec) < Québec (5.82 syll/sec). While not strong, this correlation is suggestive: between-region differences in AR might be attributed to between-regions differences in IPI length. However, age is a significant predictor of AR variation, whereas there are no age group differences in IPI length. As well, gender does not have a significant effect on AR, yet there is a significant effect of gender on IPI length. Overall, this absence of a systematic relation between, on the one hand, between-speaker differences in IPI and, on the other hand, between-speaker effects on AR indicates that there is no strong evidence in our data to support the hypothesis that IPI length can explain the effects of region, speaker age and gender on AR variation.

While our reply to the second research question is negative, the information that there is a partial correlation between average regional values in IPI length and average regional ARs points to an area of further research. Typically, each IPI unit ends with a final syllable that is longer than the syllables that precede it in that IPI unit, a process called phrase-final or pre-pausal lengthening. Speakers who have longer IPIs – such as those from Québec City – have fewer instances of this lengthening, and they would likely have a smaller number of long syllables, which would make for a faster AR. Speakers with shorter IPIs – such as those from Windsor – would likely have more instances of this lengthening, and hence slower ARs. The suggestion here is that further research should look at the nature of lengthening in IPIs as a possible source of regional differences.

3.3 Exploratory study of mean vowel durations

In a supplementary exploratory investigation – our third research question – we examined vowel durations across IPIs and ARs. Because vowel durations are highly correlated with syllable durations and because they are sensitive to processes such as final lengthening, examining the distribution of these durations may reveal sources of regional differences.

The plot in Figure 4 presents the association between IPI length (the square root of IPI) and

mean vowel duration (a measure related to AR), where the data are split into two groups by AR. The upper curve is a plot of this association for slower ARs, those less than 4.62 syll/sec (the median value); the bottom curve plots the association for faster ARs, those above the median. Both curves show that as IPI length increases, average vowel durations decrease regularly; that is, the syllables are shorter, and AR is faster. Furthermore, the curves are roughly parallel, which supports the claim that there is likely no interaction between AR and IPI length.

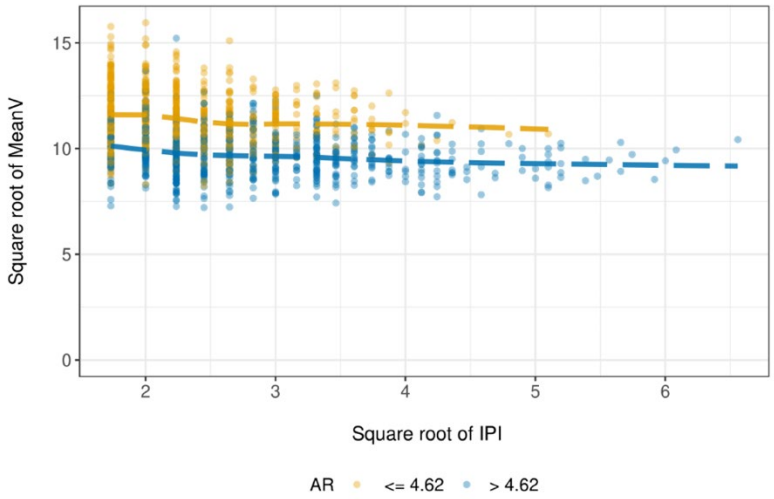


Figure 4: Relation between average vowel duration (square root of mean vowel duration) and IPI length (square root of number of syllables in IPI) for slower and faster ARs

In Figure 5, the two curves highlighted in Figure 4 are split by region. At slower ARs (the left panel), there is a noticeable difference among the regions. The curve for Québec City has a substantially different slope from the curves for Tracadie and Windsor. Specifically, for IPIs between 2 and 5 syllables (between 1.4 and 2.3 on the horizontal axis), Québec City has shorter mean vowel durations – indicating relatively faster ARs. At ARs greater than the median AR (the right panel), the slopes of the curves for the three regions are relatively similar, and the curves overlap; this suggests that these IPIs behave similarly across the three regions.

The important observation is that in the specific context of shorter IPIs with slower ARs, Québec City has noticeably shorter mean vowel durations – that is, relatively faster ARs – compared with the other regions. This suggests that potential regional differences occur in slower and shorter – as opposed to faster or longer – IPIs. This pattern can be described in terms of phrase-final lengthening. Because shorter IPIs have only a few syllables (less than 5 syllables in the left panel of Figure 5), the ARs of these IPIs are determined largely by the duration of the final syllables in the IPI. The result suggests that any regional differences are likely based on the nature of the lengthening process in each region.

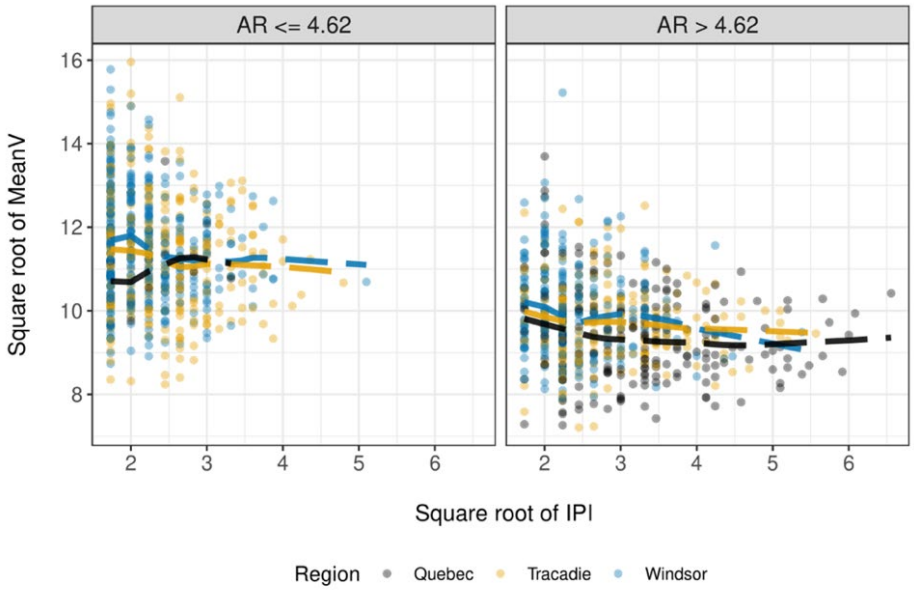


Figure 5: Relation between average vowel duration and IPI length for slower (left panel) and faster (right panel) ARs, grouped by region

4. Further discussion

4.1 Directions for future research

We highlight two of the main findings of this study: the partial but suggestive correlation between IPI length and AR for regions (in Section 3.2), and the regional differences in mean vowel durations in shorter IPIs with slower ARs (in Section 3.3). In discussing both findings we argued that the presence of lengthening at the end of the IPI unit may offer a possible explanation for regional differences in AR.

In Canadian French, it is not only the last syllable in the IPI unit that undergoes phrase-final (or pre-pausal) lengthening. The penultimate syllable can also be optionally lengthened (Walker 1984: 105-107). Furthermore, both types of lengthening can occur at the end of IPIs as well as in the Accentual Phrases that make up the IPI. This suggests the need to consider AR variation in the context of prosodic structure, whereby each IPI is a sequence of syllables that is organized at a higher level into Accentual Phrases. Our future work will look at linguistic domains such as the Accentual Phrase, with attention to the nature of accentuation and lengthening in Accentual Phrases.

This appeal to consider linguistic domains for the study of AR variation, in contrast to a phonetic domain such as IPI, has received some attention in the literature. Dankovičová (1997) measured AR in Czech in the domains of both the IPI (which she calls an interpause stretch) and the Intonation Phrase. She found that the Intonation Phrase domain is relevant for a *rallentando* (slowing down) pattern. She also observed that phrase-final lengthening tends to be greater in the final Intonation Phrase within the interpause stretch than in non-final Intonation Phrases. In his work on Texas English, Kendall (2013: 146-147) studied the relation between foot-final lengthening and phrase-final lengthening. He found that in short Intonation Phrases with one to four syllables, the overall AR is highly correlated with foot-final lengthening at the end of Intonation Phrases. As Intonation Phrases become longer, the influence of the foot-final

lengthening on AR decreases significantly. Avanzi et al. (2012a) used Accentual Phrase as the unit for measuring AR in French spoken in France and Switzerland. Two of the results that they report mirror findings based on IPI units: AR increases as the number of syllables in the Accentual Phrase increases, and there is an interaction between region and the phrase length – the Accentual Phrase length factor has a weaker effect on AR for speakers from Paris than for speakers from Switzerland, be they native or non-native francophones.

In passing, we mention that there is an interesting association between IPI length and the degree of English-French language contact, one that we have not analyzed in this paper. The regional differences in IPI length observed in this study correlate directly with the degree of English-French language contact in the three regions: IPIs are longest in Québec City, which has a low degree of contact, and shortest in Windsor, with the highest degree of contact; Tracadie, which falls between the two regions in terms of contact, has an intermediate length IPI. This observation parallels the finding in our earlier study of AR (Cichocki et al. 2024) that minority versus majority language status correlates with regional differences is AR. These parallel correlations underscore the need to explore further the nature of the hypothesized effect of language contact on temporal features such as AR and IPI length.

4.2 Comparison among regional varieties of French

We use the two temporal properties, AR and IPI length, as the basis of a brief comparison of the three Canadian varieties of French with the seven European varieties analyzed in Schwab & Avanzi (2015: 101-103). As noted above, this comparison is appropriate because data gathering for the analyses of these varieties followed the same PFC protocol. This comparison is limited to reading style. We note four points of interest.

- Average ARs have similar ranges in the European and Canadian varieties. Average reading AR in the European varieties is 5.08 syll/sec (average syllable duration is 197 msec); the Canadian average is slightly lower (4.68 syll/sec). Québec City has a considerably faster average AR (5.82 syll/sec) than the European average, while Tracadie (4.49 syll/sec) and Windsor (4.42 syll/sec) are below the European average.
- Average IPI length is longer in the European varieties (12.07 syll) than in the Canadian varieties (7.44 syll).
- Average IPI length has a narrower range among the European varieties (from 10.50 syll in Genève, to 11.33 syll in Liège, to 11.82 syll in Paris, to 12.71 syll in Lyon); the Canadian varieties have a wider range (from 5.75 syll in Windsor, to 7.81 syll in Tracadie, to 11.24 syll in Québec City).
- Unlike the seven European varieties where the effect of IPI on AR varies considerably across regions (i.e. there is a significant IPI-by-region interaction), the relationship between AR and IPI is systematic across the three Canadian regions (i.e. there is no IPI-by-region interaction).

The main differences between the Canadian and European dialects are found in IPI length and in the nature of the IPI-by-region interaction. Schwab & Avanzi (2015: 103) conclude that the regional differences in AR that they observed among the seven European regions are related to this interaction, although they state that these differences are “difficult to explain.” Our suggestion is that further work that is set in the context of linguistic domains – such as the Accentual Phrase

– may reveal new insights. A more general observation is that AR and IPI length capture significant differences and similarities among the varieties. Indeed, these temporal features might be considered as part of a general characterization of regional dialects.

5. Conclusions

This study confirms that – in reading style – region, age, and IPI length are significant predictors of AR in the three regional varieties of Canadian French studied here. The analysis makes contributions to several issues about the relationship between AR and IPI length (that is, utterance length) that appear in the literature.

- IPI length is a significant predictor of AR; the pattern – that ARs are slower in shorter IPIs and faster in longer IPIs – is consistent with patterns reported in the literature for other languages and dialects
- there is no significant IPI-by-region interaction in the three Canadian French varieties, indicating that the IPI effect on AR is systematic across regions; this contrasts with observations reported for regional varieties of European French
- IPI length does not underlie or explain differences in AR that are due to effects of region, age or gender

The exploratory study of the distribution of mean vowel durations across different ARs and different IPI lengths points to a potential IPI-by-region interaction. Regional differences appear in the context of shorter IPIs that are read at relatively slow ARs, but not elsewhere. These findings suggest that future work should consider the prosodic structure of the IPI – in terms of, for example, Accentual Phrases – and it should focus on penultimate and phrase-final lengthening.

AR and IPI length highlight similarities and differences between Canadian and European varieties of French. These temporal features can be useful in the general description and comparison of regional dialects.

Acknowledgements

The authors are grateful to the speakers from Québec City, Tracadie, and Windsor who generously agreed to participate in the *Phonologie du français contemporain* project. They also thank two anonymous reviewers for their helpful comments on an earlier version of this paper. This work was supported by the Social Sciences and Humanities Research Council of Canada (IG grant 1675-50656-10104 to S. Kaminskaïa) and by Acfas (travel grant CFC0012 to W. Cichocki).

References

Avanzi, Mathieu & Dubosson, Pauline & Schwab, Sandra. 2012a. Effects of dialectal origin on articulation rate in French. *Proceedings of Interspeech 2012*. 651-654.

Avanzi, Mathieu & Obin, Nicolas & Bardiaux, Alice & Bordal, Guri. 2012b. Speech prosody of French regional varieties. *Speech Prosody 2012*. 603–606. ISCA.

Boersma, Paul & Weenink, David. 2022. Praat: Doing Phonetics by Computer. (https://www.fon.hum.uva.nl/praat/)

Bóna, Judit. 2014. Temporal characteristics of speech: The effect of age and speech style. *The Journal of the Acoustical Society of America* 136(2). EL116-EL121.

Cichocki, Wladyslaw & Kaminskaïa, Svetlana & Hagar, Luke. 2024. Regional variation in articulation rate in French spoken in Canada. *Journal of the International Phonetic Association* 54(1). 126-145.

Clopper, Cynthia G. & Smiljanic, Rajka. 2015. Regional variation in temporal organization in American English. *Journal of Phonetics* 49. 1-15.

Dankovičová, Jana. 1997. The domain of articulation rate in Czech. *Journal of Phonetics* 25. 287-312.

Durand, Jacques & Laks, Bernard & Lyche, Chantal. 2009. Le projet PFC : une source de données primaires structurées. In Durand, Jacques & Laks, Bernard & Lyche, Chantal (eds.), *Phonologie, variation et accents du français*, 19-61. Paris: Hermès.

García-Amaya, Lorenzo & Kendro, Kelly & Henriksen, Nicholas. 2023. Regional variation, articulation rate, and pausing patterns in three varieties of Spanish. *Proceedings of the International Congress of Phonetic Sciences 2023*. 1449-1453.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press.

Grosjean, François & Deschamps, Alain. 1975. Analyse contrastive des variables temporelles de l'anglais et du français : vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica* 31. 144-184.

Jacewicz, Ewa & Fox, Robert Allen & Wei, Lai. 2010. Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America* 128(2). 839-850.

Kaminskaïa, Svetlana & Tennant, Jeff & Russell, Alexander. 2016. Prosodic rhythm in Ontario French. *Journal of French Language Studies* 26(2). 183-208.

Kendall, Tyler. 2009. *Speech rate, pause, and linguistic variation: An examination through the Sociolinguistic Archive and Analysis Project.* Durham, NC: Duke University. (Doctoral dissertation.)

Kendall, Tyler. 2013. *Speech rate, pause and sociolinguistic variation: Studies in corpus sociophonetics*. London: Palgrave Macmillan.

Malécot, André & Johnston, R. & Kizziar, P.-A. 1972. Syllabic rate and utterance length in French. *Phonetica* 26. 235-251.

Peterson, Gordon E. & Lehiste, Ilse. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32. 693-703.

Pinheiro, José & Bates, Douglas & DebRoy, Saikar & Heisterkamp, Siem & Van Willigen, Bert & Ranke, Johannes. 2017. Package 'nlme.' *Linear and nonlinear mixed effects models, version 3(1)*.

Quené, Hugo. 2008. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America* 123(2). 1104-1113.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria. (https://www.R-project.org/)

Schwab, Sandra & Avanzi, Mathieu. 2015. Regional variation and articulation rate in French. *Journal of Phonetics* 48. 96-105.

Simon, Anne-Catherine & Auchlin, Antoine & Avanzi, Mathieu & Goldman, Jean-Philippe. 2010. Les phonostyles : une description prosodique des styles de parole en français. In Abecassis,

Michaël & Ledegen, Gudrun (eds.), *Les voix des Français. En parlant, en écrivant*, 71–88. Oxford: Peter Lang.

Šturm, Pavel & Volín, Jan. 2023. Occurrences and duration of pauses in relation to speech tempo and structural organization in two speech genres. *Languages* 8(23). 1-18.

Tanner, James & Sonderegger, Morgan & Stuart-Smith, Jane & Kendall, Tyler & Mielke, Jeff & Dodsworth, Robin & Thomas, Erik. 2024. Exploring the anatomy of articulation rate in spontaneous English speech: Relationships between utterance length effects and social factors. *Proceedings of Interspeech 2024*. 467-471.

Verhoeven, Jo & DePauw, Guy & Kloots, Hanne. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47(3). 297-308.

Walker, Douglas C. 1984. *The Pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:1.27-38**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative CommonsAttribution-NonCommerciallicenseversion3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Cemeteries as semiotic landscapes: The construction of (im)migrant spaces in Canada and Germany

Grit Liebscher[1] & Doris Stolberg[2]
[1]*University of Waterloo*, [2]*Leibniz Institute for the German Language*

Abstract: This article focuses on gravestone inscriptions from cemeteries in Canada and Germany. We consider these inscriptions as publicly visible written communication and draw on frameworks from *linguistic/semiotic landscape* and *historical sociolinguistics.* Based on a qualitatively and ethnographically rooted analysis of both language and images on selected (multilingual) gravestones, and drawing on *positioning theory* and the concept of *space*, we argue that specific (im)migrant spaces and *imagined communities* are constructed through the ways in which heritage language is used and images are linked with (im)migrant places and identities. Overall, this research explores the complex interplay of socio-historical, linguistic, and semiotic aspects through which gravestone inscriptions come to index (im)migrant spaces.

Keywords: gravestone inscriptions, German, linguistic/semiotic landscape, space, positioning

Languages: English, German

1. Introduction

RESEARCH ON GRAVESTONE INSCRIPTIONS has encompassed several disciplines for many decades, including historical sociolinguistics, contact linguistics, semiotic landscape analysis, multilingualism, migration studies, and classical studies. In this article, we pursue a sociolinguistic gravestone analysis using qualitative methods that are largely informed by historical sociolinguistics and semiotic landscape analysis to analyze written language and images in the (semi-)public domain, specifically in cemeteries. Analyzing gravestone inscriptions from cemeteries in Canada and Germany, our analysis focuses on the ways in which language and other semiotic means are used to construct sociolinguistic (im)migrant spaces and imagined communities. Such inscriptions are part of publicly visible written communication that has been the focus of *semiotic landscape* research,[1] a field within sociolinguistics that has developed over the past three decades and has included the analysis of road signs, street signs, graffiti, and most recently, cemeteries.

[1] The research field started out as *linguistic landscape* (e.g. Landry & Bourhis 1997) and evolved into *semiotic landscape*, making semiotic and multimodal aspects part of this analysis (e.g. Jaworski & Thurlow 2010). While both terms are still in use, sometimes interchangeably, we have chosen *semiotic landscape*, considering our focus on images as well as language.

This article is based on a qualitative study, comparing different types of linguistic-semiotic landscaping in cemeteries. We discuss several representative examples from two different (im-)migration settings (immigration/forced migration), focusing on two cemeteries in two different countries: one cemetery is in Waterloo, Ontario, Canada, and the other one is in Flensburg, Germany. Even though neither of these cemeteries is designated specifically for (im)migrants, we will demonstrate how (im)migrant spaces are created on gravestones through linguistic and semiotic means. While multilingual linguistic resources play an important role in this, particular images on gravestones may also position the deceased within a specific (im)migrant space, often in relation to more than one language and culture (cf. Vajta 2021) and potentially in association with an imagined community.

2. Theory

From the perspective of *positioning theory* (Harré & van Langenhove 1991), it is understood that the use of language and other semiotic means (e.g. images) may position individuals and groups with regard to certain membership categories. This happens, for example, through the association of a particular dialect with a distinct region. Making use of the effects of positioning, such links to groups may result in the construction of identities (e.g. Joseph 2004) and imagined communities (Anderson 1983) in the sense of social constructivism— through a bottom-up process by human beings (in interactions, in the semiotic landscape, etc.). Positioning theory considers positioning by self or by others, and while gravestone inscriptions are commonly the result of the latter, self-positioning is possible in light of the fact that choices can be made by individuals during their lifetimes, even when carried out by others. These choices may include language and images that link to places other than the cemeteries' locations, including origins of (im)migration. Indexing such locations not only positions the deceased as (im)migrants but also creates spaces for others (e.g. the public viewer) to recognize these connections and to potentially feel included as part of an *imagined community* (Anderson 1983) as a specific kind of space. The concept of *space* has its roots in human geography and sociology (e.g. de Certeau 1988) and has since found its way into sociolinguistics as *sociolinguistic space* (e.g. Liebscher & Dailey-O'Cain 2013; Dailey-O'Cain & Liebscher 2021). Following de Certeau (1988), *space* is a constructed *place* that may rely on "enduring qualities that come to be associated with a particular place" (Dailey-O'Cain & Liebscher 2021: 13), including ideologies or ways of speaking. People also use (written) semiotic means to construct a space, for example, as an imagined community, that requires common knowledge about symbolism and language use, as the examples below will show.

3. Method

Inspired by methods of semiotic landscape and historical sociolinguistics, the research for this article originated in our interest in discovering German immigrant[2] connections on gravestone inscriptions in Canadian cemeteries. In building a database, we first visited several cemeteries in Waterloo Region[3] in 2017 and took photographs of gravestones indexing a German heritage

[2] The term *immigrant* is more appropriate than *migrant* for German immigrants who settled in Waterloo region intending to stay there permanently. We reserve the term *migrants* for our gravestone analysis in Section 4.2 that involves Germans in the European context who were forced to leave territories assigned to Poland and Russia after WWII to resettle in Germany (and other countries). Thanks to reviewer 1 for suggesting these specifications.

[3] We chose this region, since it is known in Canada for its German immigration. Grit Liebscher lives and works there and Doris Stolberg was able to visit for a research stay in 2017.

connection. In addition, we searched for relevant photographs from these cemeteries in *Find-a-grave*,[4] a database with pictures of gravestones from several cemeteries world-wide. We built a data base totalling about 300 photographed gravestones. In 2024,[5] we also visited several cemeteries in Germany and found several gravestones with (German-related) migration links. We then selected a smaller number for further qualitative analysis and will discuss representative examples in this article.

In our analysis, we draw on frameworks from semiotic landscape and historical sociolinguistics. SEMIOTIC LANDSCAPE (SL) as a strand of research within sociolinguistics has developed over the past three decades (Jaworski & Thurlow 2010). An early preference focused on a quantitative accounting of which languages occur within a particular locale (e.g. Landry & Bourhis 1997). In recent years, qualitatively focused and ethnographically rooted approaches to this research (e.g. Blommaert 2013) have become an important part of the toolkit for sociolinguistic analyses. While SL research so far has mostly focused on road signs, street signs, and graffiti, SL's framework also lends itself to analyzing cemeteries as public spaces. Drawing on Scollon & Scollon's (2003) GEOSEMIOTICS, we analyze ways by which language and other semiotic systems are located in the material world. Blommaert's (2013) ETHNOGRAPHIC LINGUISTIC LANDSCAPE ANALYSIS (or ELLA) then enables us to make sense of tombstones and epigraphs as located in time and space and with reference to and construction of people and communities. Both frameworks also draw our attention to the role of different actors in the production of gravestone inscriptions (e.g. the families, the professional carvers) and their reception (e.g. family and friends, local and non-local visitors). While most SL research was conducted in (multilingual) urban centres (e.g. Shohamy & Gorter 2008; Pennycook 2018) and minority language and migration contexts (Angermeyer 2017), cemeteries have not yet been the focus of SL research (but see Stolberg & Liebscher, under review).

HISTORICAL SOCIOLINGUISTICS investigates historical language use from different social groups and registers (e.g. Havinga & Langer 2015; Nevalainen 2015; Elspaß 2021), relating language and society from a historical perspective. Historical sociolinguistics research has contributed to a better understanding of historical multilingualism from different perspectives (Rindler Schjerve & Vetter 2003; Nevalainen 2015; Stolberg 2019). We rely on this research when contextualizing gravestone inscriptions from a perspective of social group affiliation, of migration processes, and of monolingual, multilingual, and heritage language use.

A variety of research on gravestone inscriptions has helped refine the methods for analysis. For example, works in classical and ancient studies (e.g. McDonald 2017; Clackson et al. 2020) have provided tools for linguistic contact analysis of epigraphic language in fragmentary corpora of (im)migrants' and minority languages. They emphasize the importance of investigating multiple levels of a corpus, a text's relationship to other texts of a similar type, and its physical layout and context (McDonald 2017: 45) in order to get the best informational value out of a limited data set. Our methods are also informed by research on cemeteries in other (im)migrant communities, such as those focused on long-term language contact and shift in Alsace (Vajta 2021), colonial settler contact and shift in New York (VanDam 2009) and Texas (Eckert 2002), and bilingual inscriptions in Pennsylvania (Graves 1988).

From this research we know that, linguistically, gravestone inscriptions usually rely on a ritual-formal written register. The text genre of gravestone inscriptions exhibits a clearly structured form (Jarosz 2017) that is delimited by material affordances (type of stone, available space to be written

[4] https://www.findagrave.com/ (Accessed January 9, 2025)
[5] Grit Liebscher had a research sabbatical in Germany then and we were able to visit several German cemeteries.

on, etc.) and usually obeys specific informational requirements. The underlying structure of a gravestone inscription is similar to a template, with slots that can be filled or left empty with regards to language and images. Whenever the slots are filled, decisions must be made, which results in certain kinds of positionings. This is especially the case in a setting where choices between local and immigrant languages can be made as well as images chosen that construct an (im)migrant space.

It is important to remember that gravestone inscriptions have multiple people involved in the production and reception processes. In the first instance, they are usually designed by and for a small in-group of people consisting of family and friends.[6] The production itself commonly involves professional designers and/or carvers with differential knowledge of languages and/or images. In terms of the audience or viewership of gravestones on cemeteries, they include members of the in-group as well as an out-group of visitors to a cemetery who may not know the family. After all, cemeteries are public spaces, and gravestones may be read by anyone as they are open to be viewed by the public, just as other public signs are, including posters, stickers, and graffiti.[7] Common to all these signs, including gravestones, is that they are viewed by different people who read and interpret them differently[8] and who interact with them (Fedderson et al. 2023). The result is that some viewers of gravestones may pick up different information than others. The production process may, in fact, take these audiences into account by using inscriptions that require specific knowledge (e.g. (im)migrant experience) to be recognized, as discussed further below.

All groups involved in the production process of gravestones may affect the positioning of the deceased whose names appear on gravestone inscriptions. However, in the end, it is only the deceased who are positioned a certain way, since we consider the gravestone as the final product through which these individuals are positioned in relation to certain groups such as (im)migrants that may lead to the construction of (im)migrant spaces.[9]

4. Analysis

4.1 Constructing immigrant spaces in Canada

The first four examples are taken from Parkview Cemetery, located in Waterloo, Ontario. This cemetery was established in 1966 and contains ca. 5700 graves. It is referred to as Waterloo's "main cemetery" and as "Waterloo Cemetery" on Google Maps. It is one of the three largest cemeteries in Waterloo Region. We chose to include it in our data collection, since we expected to find gravestone inscriptions indexing the long German heritage in this area. Waterloo Region has been a destination for German immigrants since the 19th century,[10] as can also be seen in gravestone inscriptions predating the ones selected for this article (Stolberg & Liebscher, under

[6] But consider gravestones of public figures such as artists and politicians, where the out-group may out-number the in-group, at least in the reception process.

[7] A difference seems to be, however, that gravestone inscriptions, while freely accessible, are usually not mainly intended for the unrelated public.

[8] Note that we left off street signs in our list of examples, since those ought to be interpreted the same by each viewer or they would not work.

[9] From a SL perspective, we don't need to know who made decisions about the inscriptions: the deceased, the family and/or friends. The professional producing the inscriptions may certainly also have an *effect* on the positioning, especially when there are errors or inconsistencies, sometimes due to limited graphic options (see Stolberg & Liebscher, under review).

[10] Though the number has been declining over the years, in 2021, 102,780, or 17.7 %, of people in the municipality of Waterloo Region still claimed German ethnic or cultural origin (Statistics Canada 2023).

review).

The first example of a gravestone constructing immigrant space can be seen in Figure 1.

Figure 1: German and English on *Albrecht* gravestone at Parkview Cemetery, Waterloo (photo: GL/DS)

The couple's names on the gravestone index a history of German(-speaking) immigration to Canada: *Albrecht* (based on its placement at the top of the gravestone, this is the couple's last name), with first names *Heinz E.* and *Erika M.* underneath. There is also a German inscription at the bottom of the gravestone: *Ruhet in Frieden* 'Rest$_{pl}$ in peace.'[11] In contrast to German, the format for the dates follows the English system, with the month before the day followed by a comma (e.g. *Feb. 17, 1920*). The marriage date in the middle also uses the English format. Overall, this mix of German and English indexes a positioning as German as well as (English-speaking) Canadian.[12]

The gravestone inscription in Figure 2 also indexes a German-Canadian positioning, in a slightly different way. In Figure 2, the last name *Ruhl* on this gravestone indexes German(-speaking) immigrant background. The dates are in German format, with day, followed by month and year (e.g. 26-05-1916). Somewhat unusual, and possibly resembling the English format, are the hyphens, since one would expect periods in the German format. This is the only element in the gravestone that might indicate an (English-)Canadian relation, aside from the grave's placement in a Canadian cemetery. A German inscription can be found at the bottom of the gravestone: *Ruhe in Frieden* 'Rest$_{sg}$ in peace,' thus again positioning the deceased as German Canadian.

[11] Both singular (sg) and plural (pl) are common with this inscription, as can also be seen in Figure 2.

[12] Note that one would not draw the same conclusion if there were, for example, Latin or (ancient) Greek, (biblical) Hebrew or Old Church Slavonic phrases on the gravestones. Those represent a different kind of cultural phenomenon, not indicating active bilingualism or associating people with nation-states as the living languages in our examples would. Thanks to reviewer 2 for pointing this out.

Figure 2: German (and English) on *Ruhl* gravestone at Parkview Cemetery, Waterloo (photo: GL/DS)

Figure 3 shows us yet another German Canadian positioning. Here again, both first and last names index a German(-speaking) background. Since we only find the years of death but not days and months, there is no German or English positioning with the dates, as in Figures 1 and 2 above. In fact, the choice to put only the years (intentionally or unintentionally) avoids a positioning as German or Canadian. German is used, however, in *Geb. Kaufmann* 'née Kaufmann' but English is then used for *Rest in peace*, a mix that positions the deceased as both German and (English-speaking) Canadian.

Figure 3: German and English on *Grimmecke* gravestone at Parkview Cemetery, Waterloo (photo: GL/DS)

The last gravestone in this section, Figure 4, is entirely in German. The gravestone in Figure 4 has the last name on top (*Aysslinger*), right above a heart-shaped ornament that has a southern German, Austrian, or Swiss touch, since it is a style reminiscent of folkloric images in these areas. Below the ornament, we find the names, that of the deceased (*Utz*) as well as his partner (*Karin*), whose year of death is not written on the gravestone, which indexes that she was still alive at the time when the photo was taken. German abbreviations are used with years of birth and death: *Geb.* 'born' and *Gest.* 'died,' and an inscription at the bottom is also in German: *Bis in alle Ewigkeit* 'For all eternity.' There is nothing English about this gravestone, except its placement in a Canadian cemetery. The language used on the gravestone, which is entirely German, positions them with a strong connection outside of Canada, constructing a German(-speaking) European space in Canada.

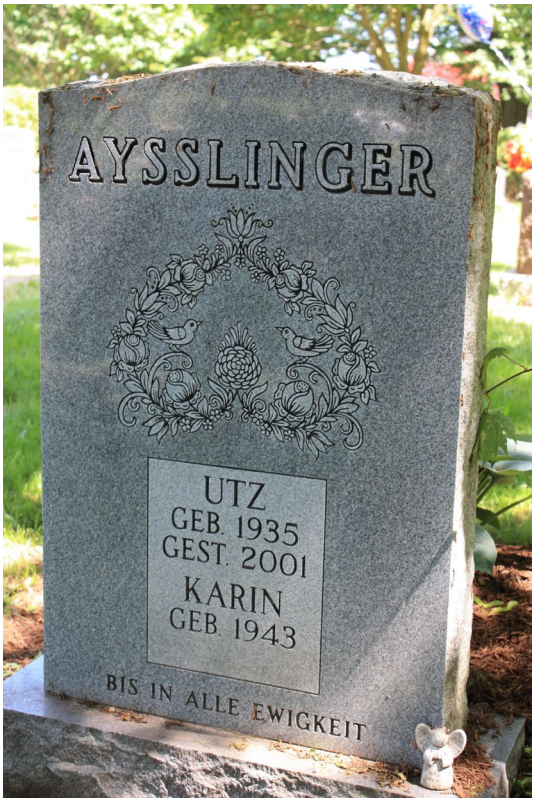


Figure 4: German on *Aysslinger* gravestone at Parkview Cemetery, Waterloo (photo: GL/DS)

4.2 Constructing migrant spaces in Germany

The examples in this section, showing constructions of migrant spaces in Germany, come from Adelby cemetery in Flensburg at the German-Danish border in Germany's north-western region.

The gravestone in Figure 5 below is especially striking for the image engraved on it.

Figure 5: *Spang* gravestone with East Prussia map at Adelby Cemetery, Flensburg (photo: GL/DS)

The gravestone in Figure 5 contains the names for the deceased: *Edgar Spang* and *Anne-Marie Spang*, presumably husband and wife, since the wife's maiden name (*Stünitz*) is provided as well. The fact that the names are engraved in a hand-written fashion gives this gravestone a personal touch, constructing a less formal and less distanced relationship to the viewer. The names as well as the dates of birth and death in the typically German format position the deceased as German. However, the image in the upper part of this gravestone adds another dimension to it. After an internet search for similar images, the image was recognizable to us as a map of East Prussia,[13] an area in today's Poland and Russia that was part of Germany before WWII. Placing the map on a gravestone in Germany positions the deceased as German migrants from this area, since they were born in 1911 and 1917, respectively. In fact, the map creates an imagined community for those who recognize it and find themselves connected to this group. The image of the map can potentially be seen as a political statement, when associating it with the loss of East Prussia to Poland and Russia. However, the intention may simply have been to index a place of migration for the deceased, which can also be done through providing the name for a place of birth, as can be seen on the next gravestone in Figure 6.

[13] https://images.app.goo.gl/3SjJWmqRSrANYSy98 (Accessed October 24, 2024)

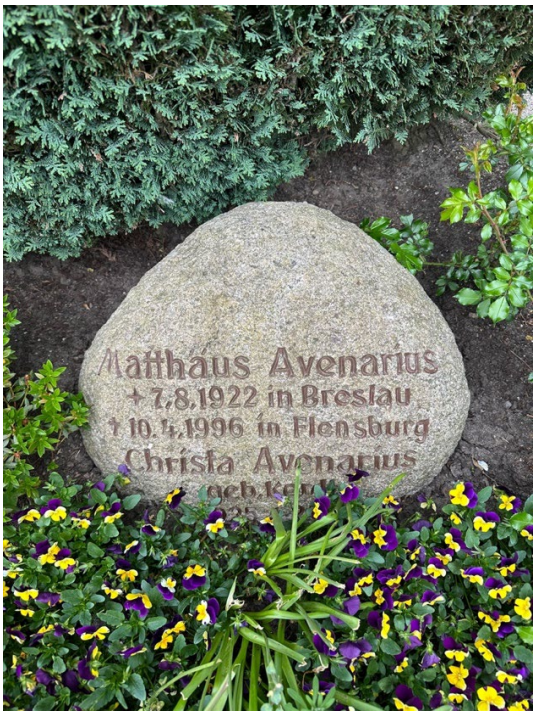


Figure 6: Migration marker on *Avenarius* gravestone at Adelby Cemetery, Flensburg (photo: GL/DS)

The gravestone in Figure 6 contains the couple's names: *Matthäus Avenarius* and *Christa Avenarius*, together with the wife's maiden name. In addition, the dates and places of birth and death appear on the gravestone.[14] Notably, the place of birth for *Matthäus Avenarius* is named as *Breslau*, the old name for today's Wrocław in Poland, before this previously German area became part of Poland after 1945. Like the map in Figure 5, the use of the old name makes this history relevant, including the forced migration of the German population from this area. Again, this positions the deceased as part of this group and allows viewers to identify as part of the imagined community of forced migrants (from this area). The difference in providing the name of the place, compared to an image, lies in the fact that the name can be read by anyone, but an image may only gain immediate recognition by someone who knows. The image, then, is designed for a specific group of people who are in the know. Another example for an image indexing migrant experiences can be found in our last example, in Figure 7.

[14] They are visible for the husband, though the wife's dates are partially obstructed by the flowers.

Figure 7: *Meeske* gravestone with Danzig icon at Adelby Cemetery, Flensburg (photo: GL/DS)

Similar to the previous two gravestones, the one in Figure 7 shows the names of a couple, with dates of birth from the same decade as those of the deceased on the *Spang* gravestone (Figure 5). In addition, there is an image in the upper left corner. The image shows a crown with two crosses underneath. This symbol is part of today's Gdańsk flag,[15] likely going back to a medieval symbol. It has also been used as the symbol during the "Free City of Danzig" between 1923 and 1939 (e.g. on its coin, the Danzig Gulden).[16] Gdańsk is again in an area that was part of Germany (and Prussia) before 1920 and during WWII and was assigned to Poland after 1945. As with the other two examples in this section, the image positions the deceased as migrants from a formerly German area, affected by the forced migration after 1945.

5. Conclusion

In this article, we focused on selected gravestone inscriptions from two cemeteries, one in Canada and one in Germany. Using qualitative approaches from semiotic landscape and historical linguistics, we argued that specific (im)migrant spaces and imagined communities were constructed through the use of language and images on these gravestones.[17]

In the first part of the analysis section with a focus on a cemetery in Canada, we discussed the construction of German Canadian sociolinguistic spaces through the mix of German as well as English language use, including formatting such as with dates. Through these semiotic means, the deceased positioned themselves and/or were positioned as both German and (English-speaking) Canadian. The fact that these gravestones are located in Canada makes a German heritage connection relevant (e.g. through language use of the European birthplace). By pointing to the origin of migration while at the same time linking to localized spaces and positionings through the use of English, a sociolinguistic German heritage space in Canada and an imagined community of German(-speaking) immigrants in Canada are constructed.

[15] https://en.wikipedia.org/wiki/Gdańsk (Accessed October 23, 2024)
[16] https://en.wikipedia.org/wiki/Danzig_gulden (Accessed October 23, 2024)
[17] Another possibility for positioning may be done through the selected gravestone material. Vaijta (2021: 27), for example, discussed how Alsatian sandstone was used to point to Alsace as the place of origin.

The gravestones located in Germany also used language and images to index the deceased's place of origin, leading to a positioning as a migrant. More specifically, we argued that the link to the history of forced migration creates opportunities for viewers with similar experiences to recognize these places and to see themselves as part of this imagined community. In contrast to the Canadian setting, language choice could not indicate origin in the German cases because the pre- and post-migration (written) language was the same. Therefore, different means had to be applied, such as visuals and explicit naming of places, equally succeeding in indicating a migrant history.

Overall, language and images on gravestone inscriptions contribute to positionings of the individuals whose names appear on them. At the same time, they afford the construction of sociolinguistic spaces and may make these spaces recognizable for others to include themselves as part of imagined communities. Notably, compared to other, more ephemeral positioning events, inscriptions on gravestones may last for decades or centuries, keeping history alive.

Acknowledgments

We would like to thank the Ontario/Baden-Württemberg Faculty Mobility Programme for supporting Doris Stolberg's field work in Waterloo. We also appreciate the editors' work and the very helpful comments by two anonymous reviewers.

References

Anderson, Benedict. 1983. *Imagined community: Reflections on the origin and spread of nationalism*. London: Verso.

Angermeyer, Philipp S. 2017. Controlling Roma refugees with 'Google-Hungarian': Indexing deviance, contempt, and belonging in Toronto's linguistic landscape. *Language in Society* 46(2). 159-183.

Blommaert, Jan. 2013. *Ethnography, superdiversity and linguistic landscapes.* Bristol: Multilingual Matters.

Clackson, James & James, Patrick & McDonald, Katherine & Tagliapietra, Livia & Zair, Nicholas (eds.). 2020. *Migration, mobility and language contact in and around the ancient mediterranean.* Cambridge: Cambridge University Press.

Dailey-O'Cain, Jennifer & Liebscher, Grit. 2021. Constructing the motherland: German-Canadian positioning and the tensions between place and space. In Nedashkivska, Alla & Kuße, Holger (eds.), *Boundaries and belonging: Language, diaspora, and motherland (Journal of Belonging, Identity, Language, and Diversity (J-BILD*)) 5(1). 11-31.

de Certeau, Michel. 1988. *The practice of everyday life.* Berkley*:* University of California Press.

Eckert, Eva. 2002. Language variation, contact, and shift in tombstone inscriptions. In Janda, Laura A. & Franks, Steven & Feldstein, Ronald (eds.), *Where one's tongue rules well: A festschrift for Charles Townsend* (Indiana Slavic Studies 13), 193-212. Bloomington: Indiana University.

Elspaß, Stephan. 2021. Language standardization in a view 'from below.' In Ayres-Bennett, Wendy & Bellamy, John (eds.), *The Cambridge handbook of language standardization,* 93-114. Cambridge: Cambridge University Press.

Feddersen, Richard & Liebscher, Grit & Dailey-O'Cain, Jennifer. 2024. Turn-taking in the interactive linguistic landscape. *Linguistic Landscape* 10. 22-54.

Graves, Thomas. 1988. Pennsylvania German gravestones: An introduction. *Markers* 5. 60-95.

Harré, Ron & van Langenhove, Luk. 1991. Varieties of positioning. *Journal for the Theory of Social Behaviour* 21. 393-407.

Havinga, Anna & Langer, Nils (eds.). 2015. *Invisible languages in the nineteenth century.* Oxford: Peter Lang.

Jarosz, Józef. 2017. *Grabinschrift - eine Textsorte im Wandel: eine diachrone Studie am deutschen epigrafischen Material 1780-2015* [Funerary inscription - a changing text type: a diachronic study of German epigraphic material 1780-2015]. Dresden: Neisse Verlag.

Jaworski, Adam & Thurlow, Crispin (eds.). 2010. *Semiotic landscapes: Language, image, space*. New York: Bloomsbury Publishing.

Joseph, John E. 2004. *Language and identity: National, ethnic, religious*. Basingstoke: Palgrave Macmillan.

Landry, Rodrigue & Bourhis, Richard Y. 1997. Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology* 16(1). 23-49.

Liebscher, Grit & Dailey-O'Cain, Jennifer. 2013. *Language, space, and identity in migration.* Basingstoke: Palgrave Macmillan.

McDonald, Katherine. 2017. Fragmentary ancient languages as "bad data": Towards a methodology for investigating multilingualism in epigraphic sources. *Sociolinguistica* 31(1). 39-56.

Nevalainen, Terttu. 2015. What are historical sociolinguistics? *Journal of Historical Sociolinguistics* 1(2). 243-269.

Pennycook, Alastair. 2018. Linguistic landscapes and semiotic assemblages. In Pütz, Martin & Mundt, Neele (eds.), *Expanding the linguistic landscape: Linguistic diversity, multimodality and the use of space as a semiotic resource*, 75-88. Bristol: Multilingual Matters.

Rindler Schjerve, Rosita & Vetter, Eva. 2003. Chapter 2. Historical sociolinguistics and multilingualism: Theoretical and methodological issues in the development of a multifunctional framework. In Rindler Schjerve, Rosita (ed.), *Diglossia and power: Language policies and practice in the 19th century Habsburg Empire*, 35-68. Berlin: De Gruyter Mouton.

Scollon, Ron & Scollon, Suzie Wong. 2003. *Discourses in place: Language and the material world.* London: Routledge.

Shohamy, Elana & Gorter, Durk (eds.). 2008. *Linguistic landscape: Expanding the scenery.* London: Routledge.

Stolberg, Doris. 2019. Canadian heritage German across three generations: A diary-based study of language shift in action. *Journal of Historical Sociolinguistics* 5(2). 2-35.

Stolberg, Doris & Liebscher, Grit. Under review. Geb. ('Born') in Berlin, Germany: A semiotic construction of German heritage in Canadian cemeteries. In Brown, Joshua R. & Kühl, Karoline (eds.), *Gravestones of multilingual communities as sources for historical sociolinguistics*. (*Journal of Historical Sociolinguistics*, special issue).

Vajta, Katharina. 2021. Identity beyond death: Messages and meanings in Alsatian cemeteries. *Mortality* 26(1). 17-35.

VanDam, Kara. 2009. Dutch-American language shift: Evidence from the grave. *LACUS Forum* 34. 31-40.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:1.40-52 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Some demolinguistic trends of the Francophone minority in Canada in the 21st century

Jocelyn van Leeuwen
University of Waterloo

Abstract: French-speaking Canadians outside of Quebec reside in a minority situation where they often cannot live only in French. This environment impacts the linguistic vitality of the group as some choose to speak the language of majority, English, both in public and at home. This study offers a descriptive analysis of mother tongue, home language, and language shift trends of the minority Francophone population gathered from the 2001 to 2021 census data. The study found that the francophone communities outside of Quebec saw a continuing decrease for all three language use parameters. In addition, linguistic family dynamics and language transmission were considered. A closer examination showed that over 80% of children born to two Francophone parents had French as their mother tongue, whereas an average of about 30% of children born to only one Francophone parent had French as their mother tongue in 2021. This indicates a considerably lower French transmission for children born to exogamous families. Overall, a lower proportion of Francophones in a region appears to negatively affect the transmission of French in English-speaking Canada.

Keywords: Francophone minority, census, demolinguistics, exogamy, endogamy

Languages: English, French

1. Introduction

1.1 Canada's linguistic landscape

THE FOUNDATION OF CANADA was built upon its colonization by both French and British settlers and their back-and-forth battle for the control of the territory. Ultimately, the British prevailed; however, the struggle for Canada remains etched in its history forever with English and French as Canada's two official languages. Since the Confederation of the country in 1867, English and French have equal status on the federal level (Sheppard 1971). Despite having equal status, services and institutions in both languages are not available everywhere throughout the country. This is in part due to each of Canada's 10 provinces and three territories having their own official language(s). Quebec is the only province with French as its official language, New Brunswick the only bilingual province, and the remaining provinces are officially English-only.

Outside Quebec, the French-speaking population is the linguistic minority, and these speakers are the focus of our study. We acknowledge that speakers of Indigenous languages and ethnic languages are also part of the linguistic minority; however, we will focus solely on the dynamics

between the two official language groups. Since the Francophone population in the three territories is not large enough to affect our results, with fewer than 5000 Francophones (unilingual, bilingual or multilingual) in the 2021 census data and a significant number of Indigenous speakers, we have decided not to include any of the territories in our analysis.

In this article, we define a Francophone as a person whose mother tongue is French and an Anglophone as a person whose mother tongue is English, where one's mother tongue, as defined by Statistics Canada, is the language first learned in childhood and still understood today. Those with a mother tongue other than English and French are referred to as Allophones. There are also people who learned two or more languages at the same time, and who fall under the definition of bilingual, trilingual speakers, etc. For the purposes of this article, we will also define a French user as someone whose home language (the language most spoken at home, as defined by Statistics Canada) is French. Although Francophones are often French users and vice versa, these identifiers are independent from each other.

The purpose of our study is to examine demolinguistic dynamics from 2001 to 2021 of French mother tongue and home language, as well as language shift, endogamy/exogamy and language transmission trends. This will provide a comprehensive 20-year overview of the minority French community in Canada which has not been done since the 2006 census.

1.2 Languages in contact and demolinguistic dynamics

Languages in contact do not always have equal status. This can lead to the minoritization of one of the speaker groups, which has consequences (Hambye 2019). For instance, if the Francophone population in English-speaking Canada is not large enough or not concentrated in a specific area, some services in the minority language may not be available (Cartwright 1988). In some cases, this means that even schooling and medical services are not available in French and Francophones must communicate in English out of necessity.

Besides minoritization, a weak demographic presence of minority speakers in a region can greatly affect the linguistic vitality of the community (Remysen 2003). Our definition of linguistic vitality is adopted from that of Giles et al. (1977: 308) where it is defined as a language community's behaviour to act as a distinctive and active group when that language is in contact with another. While demographic presence of a language is at the core of linguistic vitality (Landry & Allard 1994; Jedwab 2002; Corbeil 2005), the latter also depends on the language's status and institutional support (Giles et al. 1977: 309; Mougeon 2014). They contribute to the perception of the language as being valued and encourages its use, thus strengthening the linguistic vitality of a group (Marquis et al. 2022: 4)

The opportunity to speak the minority language decreases when the demography of a minority linguistic group is reduced. This can influence an individual's choice to speak the minority language in their private life and as parents, pass the language down to their children. It can also affect the quality of a language (e.g. Mougeon 2004) or an individual's feeling of insecurity towards it (Bergeron et al. 2022) because of this or because of its lesser status. When a group has a weak linguistic vitality, we start to see a language shift which "is the change in the habitual use of one language to that of another." (De Vries 1994: 62)

The linguistic dynamics of a minority population are thus affected both quantitatively (e.g. birth rate, mother tongue, language use, language transmission) and qualitatively (e.g. linguistic insecurity and linguistic variation). The focus of our study is on some quantitative aspects of French language dynamics as collected by the Canadian census.

1.3 The Canadian census

The Canadian census is administered every five years by the national statistics office, Statistics Canada, and it collects demographic, social, and economic data about the Canadian population. Depending on the census year (ending with "01" or "06"), there are two census questionnaires for the general population, a long version and a short version. The short version of the questionnaire is administered to 80% or 75% of the population, depending on the census year, and a long version to 20% or 25% of the population. Both versions contain the same questions relating to demographics and language. We will consider that Statistics Canada has properly selected a representative population for the long questionnaire, and we will also treat the responses from both the short and the long questionnaire as equal. The census questionnaires are available on the Statistics Canada website (https://www.statcan.gc.ca/).

The census began first collecting data on one's mother tongue in 1901, and on the language most frequently used at home in 1971 (Statistics Canada 2021-2022). Since 2001, the questionnaires have asked about all languages spoken at home, the language most spoken at home and other languages spoken on a regular basis. For our analysis of the home language, we considered only the language(s) most spoken at home. This information provides the data for language shift calculation (see Castonguay 2001 among others). By crossing answers to language questions with respondents' demographic information, we can look at the language(s) each family member has as their mother tongue. This allows us to examine language transmission. Besides this, we can also determine if a couple is linguistically endogamic (i.e. each person in the couple has the same mother tongue) or if they are linguistically exogamic (mixed) (i.e. each person in the couple has a different mother tongue). Because this study focuses on the minority Canadian Francophone population, we will consider a couple as endogamous if both partners are Francophone, and as exogamous, couples where one person is a Francophone, and the other has a mother tongue other than French. Studying couples that have at least one Francophone will allow us to observe trends regarding the Francophone population. We are not concerned about the behaviour of Anglophones and Allophones other than their involvement in an exogamous relationship.

2. Literature review

There have been many demolinguistic studies on the presence of minority Francophone communities based on Canadian census data (e.g. De Vries 1994; Mougeon & Beniak 1994; Langlois 2000; Castonguay 2001). When the census began inquiring on the population's home language in 1971, it provided further insights into language patterns. Past research that focused on Anglophone regions using census data from 1971 to 2001 observed a decline in the proportion of Francophones for both mother tongue and home language statistics as well as an increase in language shift (also called anglicization or assimilation) (Castonguay 1993, 1997, 2001, 2005; Langlois 2000). On the other hand, the number of people who could speak both English and French in the Francophone population increased over multiple decades. This is an important observation because bilingualism in a minority setting often leads to language shift (Mougeon & Beniak 1994). As a result, bilingualism is considered subtractive in a minority setting since the speaker loses some mastery of their mother tongue to account for the command of the language of majority (De Vries 1994; Landry & Allard 1994; Bernard 1997).

To account for the observed tendencies, Castonguay (2001) examined the importance of the density of a minority Francophone population to maintain one's mother tongue. He did this by

comparing the demographics of Francophones in New Brunswick with those in Ontario as well as the evolving trends of the rate of anglicization from 1971 to 2001. He found that in regions with a higher density of Francophones —in New Brunswick and in the North and East regions of Ontario —there was a lower rate of language shift to English among Francophones compared to the regions with a lower Francophone density. Castonguay (2005) and De Vries (1994) also found that, overall, language shift was lower in New Brunswick compared to Ontario.

Not only does the language of the community matter for a language's vitality, but so does the linguistic environment at home. When couples are linguistically mixed, it can have a large effect on the language or languages their children learn. In a minority situation, if one parent speaks French but the other has a different native language but knows or speaks it, it is more likely that the child will speak French (Bouchard-Coulombe 2011; Vézina & Houle 2014). Furthermore, if the mother is the minority language speaker, the child is more likely to speak the minority language (Bouchard-Coulombe 2011).

Finally, the factor that has a direct impact on maintaining the French language within a community is fertility rate. It is generally considered that to maintain a population, there must be an average of 2.1 children born per mother, or a fertility rate of 2.1 (Espenshade et al. 2003). Presently, this threshold is not being met as the fertility rate is decreasing rapidly among first-world countries and Canada is no exception with fertility rates generally below 1.6. According to Provencher & Galbraith (2024), the average fertility rates in Canada are as follows: 1.54 (2001), 1.63 (2006 and 2011), 1.59 (2016), and 1.4 (2021). This trend leads us to think that fertility rate directly affects the demographic vitality of the Francophone population because even if there were no other factors that affected the intergenerational transmission of French, the Francophone community may still not be producing enough children to replace themselves.

In the past, researchers have used the rate of language shift/maintenance (where the number of native speakers is compared to the number of people using that language at home) as the main measure of linguistic vitality (De Vries 1994; Castonguay 2001, 2005, 2021). For example, Castonguay (2021) looked at data relating to home language and mother tongue for the minority Francophone population between 1971 and 2016. He was able to paint a worrying picture – a significant decrease of Francophones and French users combined with a high rate of language shift – and concluded that the vitality of the Francophonie in Canada is in an extremely precarious situation. De Vries (1994) used a similar approach and concluded that the Francophonie in minority settings does not have an optimistic outlook on their long term "survival," especially for those with a very small Francophone population, such as those in British Columbia and Newfoundland, where about half of the Francophones were English users at home according to the 1986 census data.

Alternatively, Mougeon & Beniak (1994) adopted a more nuanced perspective where they considered other factors susceptible to affect the linguistic vitality of the minority Francophone population than the choice of one's home language. This included bilingualism, linguistic exogamy, and language use in the community and at school by supplementing census data with their own data from previous studies such as Mougeon & Beniak (1989). Mougeon (2014) is another example of a nuanced study on linguistic vitality where the author looked at linguistic exogamy, intergenerational linguistic transmission, the use of French in the workplace, and attending French-language schools in addition to mother tongue, home language, and language shift. From this approach, Mougeon stated that "on a conclu, peut-être trop rapidement, à la disparition future des minorités francophones hors Québec." (p. 211) Mougeon's findings align with Langlois' (2000) observations that while language shift (which considers language use only

in the familial sphere) can be used as a benchmark of the linguistic vitality of a linguistic community, language use extends beyond the family and into the workplace and other social interactions.

As we saw, previous studies focus on mother tongue and home language between 1971 and 2006 censuses (Mougeon 2014; Castonguay 2021). Very few of them focus on multiple factors related to a community's linguistic vitality (e.g. Mougeon 2014), and some studies narrowed in on certain regions to compare the effect of regions with more Francophones and those with a smaller Francophone population (e.g. Castonguay 2001, 2005). What is missing is a recent study on multiple census years while also investigating different factors that could lead to changes in the linguistic vitality of the minority francophone population. This paper will examine the three measures of language use (mother tongue, home language, and language shift) from the 2001 census to the 2021 census in addition to investigating other variables that may contribute to the changes in these language use parameters – exogamy and its role in intergenerational transmission following Mougeon (2014) to provide a broader picture of the minority Francophone population. We thus contribute a demolinguistic perspective of the minority Francophone population in the 21st century in Canada.

Our study focuses on four regions. The first region is Canada as a whole; many researchers studied Canada previously, which allows us to compare our data on Canada with previous findings. The other three regions examined in this article are the provinces outside Quebec: New Brunswick, Ontario and the ensemble of the remaining seven Anglophone provinces – British Columbia, Alberta, Saskatchewan, Manitoba, Nova Scotia, Prince Edward Island, and Newfoundland and Labrador. We will hereafter refer to these seven provinces as the *Seven Provinces*. New Brunswick and Ontario are considered individually because, historically, outside of Quebec, New Brunswick has the largest percentage of Francophones in the country at around 30% of its population, while Ontario has the largest number of Francophones in the country – 524690. Despite being the largest Francophone population outside of Quebec, the Franco-Ontarian population does not lately surpass 4% of the total of Ontario's population. The Seven Provinces have a much smaller Francophone population generally with around 1% of their population and approximately a quarter million people total.

In this context, our study aims to answer the following questions:

(1) Over the period 2001 to 2021, have there been changes in the demographic presence of Francophones in the four regions under study? If yes, what is the nature of these changes?
(2) What were the language shift trends in the regions of interest between 2001 and 2021?
(3) Do regions with a higher density of Francophones show a stronger linguistic vitality?
(4) What role does exogamy play in the role of linguistic vitality and language transmission?

Our study provides a 20-year depiction of the language trends of the minority Francophone population in Anglophone Canada for the years 2001 to 2021 inclusive by taking into consideration the speakers' demographics and language data from the census questionnaires.

2. Methodology

To answer research questions, we undertook a descriptive study of demolinguistic dynamics based on data sourced from the Statistics Canada website for the 2001, 2006, 2011, 2016, and 2021 censuses. We used demographic information related to age and place of residence and associated

it with responses to the language questions in the Canadian census.

We will first follow Castonguay's approach in his 2001 article and share our observations on the three measures of language use: mother tongue and home language, and language shift. We will then follow Mougeon's (2014) approach by considering linguistic exogamy and its possible effect on language transmission and language use.

We are treating all responses to the census in the same manner, even though the questions that ask about mother tongue and home language have varied slightly between 2001 and 2021, as well as the order in which the questions are asked. While it is possible that the order the questions are asked in may have affected a person's answer (i.e. selecting one or multiple languages), this is beyond the scope of this research.

For the purposes of our analysis, we consider Francophones to include unilingual French speakers (where the individual has French as their sole mother tongue) as well as bilingual and multilingual Francophones where the individual has French as their mother tongue together with a second or multiple others.

To calculate the size of the Francophone population, we used Castonguay's (2021) equation:

(1) $Total\ number\ of\ Francophones = MT_{FR} + (MT_{FR+EN})/2 + (MT_{FR+NO})/2 + (MT_{FR+EN+NO})/3$

where MT is the mother tongue with FR designating French as the mother tongue, EN designating English as the mother tongue, and NO signifying a non-official language.

Hence, the total number of Francophones is calculated as the sum of people who have one mother tongue (French), half the number of bilingual Francophones (who have two mother tongues: French and either English or a non-official language), and a third of people who have three mother tongues, French and two other languages (English and a non-official language or two non-official languages).

For trends related to mother tongue, home language and language shift, we will use the above equation. For the analysis of linguistic exogamy, we will differentiate among unilingual, bilingual, and/or multilingual Francophones and will provide the necessary details in Sections 3.4 and 3.5. Additional formulae calculating percentage change, language shift, and percentage of endogamous or exogamous couples in a region are defined in Section 3.

3. Results

In recent years, Canada has begun to see a decrease in the percentage of native speakers for both English and French, largely in part due to immigration where most of the people coming into the country are Allophones (Castonguay 2021). In the following subsections, we will look at language trends for the four target regions to provide a broad overview of the changing dynamics of the Francophone population over a span of 20 years. We would like to recall that the results for Canada include Quebec.

3.1 French mother tongue

Table 1 below shows the number of Francophones in 2001 and 2021 in the four focus regions as well as the percentage of the population. The table also includes the difference in percentage between 2001 and 2021, noted as Diff(%). We also normalized Diff(%) relative to the 2001 values to get a better idea of the change (noted as %change) that this difference represents with respect to

the 2001 situation. Diff(%) was calculated by subtracting the % value of 2001 from 2021, and %change was calculated with the following equation:

(2) *Percentage change (%change)* = $[(P_{2021}-P_{2001})/P_{2001}]$ x 100%
where P corresponds to the percentage of the population of the year denoted in the subscript.

The percentage change is calculated as the percentage of the population in 2021 less the percentage of the population in 2001 divided by the percentage of the population in 2001 times 100.

Table 1: French mother tongue for the 2001 and 2021 Canadian censuses for Canada, the Seven Provinces, New Brunswick, and Ontario

Year	2001 (N)	2021 (N)	2001 (%)	2021 (%)	Diff(%)	%change
Canada	6782294	7411489	22.88	20.24	-2.64	-11.54
Seven Provinces	229309	239158	1.02	0.85	-0.81	-16.67
New Brunswick	239357	232411	33.26	30.40	-2.86	-8.60
Ontario	509264	524690	4.51	3.72	-0.79	-17.52

We observe that from 2001 to 2021, the number of Francophones increases for Canada (6782294 to 7411489), the Seven Provinces (229309 to 239158), and Ontario (509264 to 524690), and decreases for New Brunswick (239357 to 232411). However, in all four regions, the percentage of Francophones decreases over the 20 years (note negative values for Diff(%)). In the final column, we find that while Diff(%) values for Canada (-2.64%) and New Brunswick (-2.86%) seem greater than those for Ontario (-0.79%) and the Seven Provinces (-0.81%) from 2001 to 2021, their normalized values (-11.54% and -8.60%, respectively) are surpassed by those of Ontario (-17.52%) and the Seven Provinces (-16.67%). This indicates that Francophone population is decreasing more rapidly in Ontario and the Seven Provinces (the two regions with a lower concentration of Francophones, 3.72% and 0.85% in 2021 respectively in Table 1), compared to New Brunswick and Canada who have a higher density of Francophones (30.40% and 20.24% in 2021 respectively in Table 1).

3.2 French home language

Similarly to Table 1, Table 2 details the number and the percentages of users of French as home language in the four targeted regions in 2001 and 2021.

Table 2: French home language for the 2001 and 2021 Canadian censuses for Canada, the Seven Provinces, New Brunswick, and Ontario

Year	2001 (N)	2021 (N)	2001 (%)	2021 (%)	Diff (%)	%change
Canada	6531375	7242028	22.04	19.78	-2.26	-10.25
Seven Provinces	86869	88164	0.83	0.67	-0.16	-19.28
New Brunswick	217773	206539	30.26	27.01	-3.25	-10.74
Ontario	307297	292029	2.72	2.07	-0.65	-23.90

Again, we see an increase in the number of French language users for Canada (710653 speakers), and the Seven Provinces (1295 speakers) but a decrease in the number of French users in Ontario (-15268 speakers) and in New Brunswick (-11234 speakers). Across the board there is a decrease in the percentage of French users (Canada, -2.26%; Seven Provinces, -0.16%; New Brunswick, -3.25%; Ontario, -0.65%). The percentage change is more marked for the Seven Provinces and Ontario (-19.28% and -23.90% respectively) which are the extreme minorities (less than 4% of the total population) compared to Canada and New Brunswick (-10.25% and -10.74%, respectively) where the minority Francophone population is over 20% of total population.

Comparing the values from Table 1 to Table 2, we note that there is a smaller percentage and a smaller number of people with French as the home language (Table 2), compared to those with French as their mother tongue (Table 1) across all four regions. This leads us to Section 3.3 where we measure the language shift by using the mother tongue and home language data values.

3.3 Language shift rate

The language shift rate is measured based on the mother tongue and the home language data, as in Castonguay (2001) (3):

(3) $\textit{Language shift} = [(\textit{Nmb of people}_{MT} - \textit{Nmb of people}_{HL})/\textit{Nmb of people}_{MT}] \times 100\%$

The language shift is calculated as the difference between the number of people with mother tongue (MT) X and the number of people with home language (HL) X divided by the number of people with mother tongue X and multiplied by 100. The result provides a snapshot of what percentage of the linguistic population of language X no longer regularly speaks the language X anymore each census year.

If the language shift equals zero, this means that the number of people with mother tongue X is equal to the number of people with home language X and the population of language X has not lost any users. If the language shift is negative, this means that there are more users of language X than the number of native speakers. This can occur when people with a different mother tongue have chosen to use language X at home. On the other hand, if the language shift is positive, this indicates that there are fewer people who use language X than the number of native speakers and that some of the native speakers have chosen to use a different language at home. A higher language shift value indicates a higher number of people who have abandoned language X at home. Recall that both mother tongue and home language values in this study account for all unilingual, bilingual, and multilingual Francophones or French users.

Figure 1 below charts the language shift dynamics from 2001 to 2021 for the four regions of this study. There is a noticeable difference in language shift between Canada and New Brunswick, on the one hand, and between Ontario and the Seven Provinces, on the other hand. These trends are in reverse correspondence with the number of Francophones in Table 1.

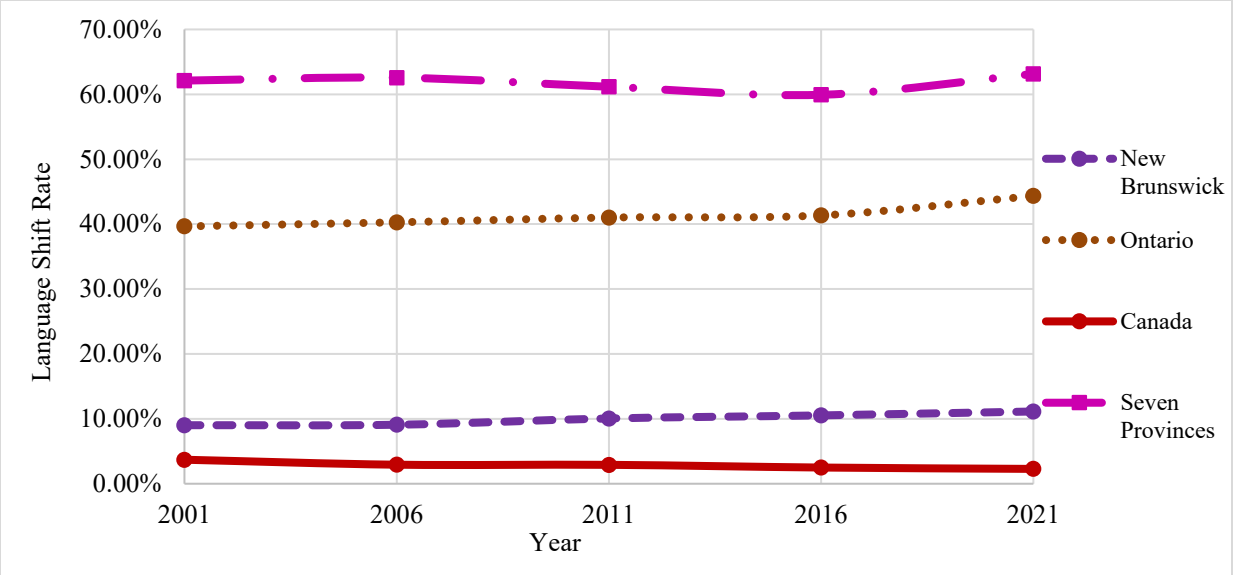


Figure 1: Language shift for the 2001 to 2021 Canadian censuses for Canada, the Seven Provinces, New Brunswick, and Ontario

Interestingly, while we saw a decrease in the percentage of Francophones and French users in Sections 3.1 and 3.2, we see a decreasing trend for language shift for Canada from 3.70% to 2.29%. On the other hand, we see an increasing trend for language shift in New Brunswick (9.02% to 11.13%), Ontario (39.66% to 44.34%) and the Seven Provinces (62.12% to 63.14%). With the language shift trending in a different direction for Canada compared to the Anglophones provinces, we suspect that decreasing language shift in Canada can mainly be accounted for by the Francophone Quebecois population.

These numbers paint an alarming picture for the state of extreme minority French populations in the Anglophone provinces. By examining the exogamy/endogamy data, we are looking to understand what may be contributing to these changes.

3.4 Linguistic exogamy

In this section, we are only considering unilingual adult Francophones for ease of calculations. The percentage of couples were calculated using language data from Statistics Canada for endogamous couples as in (4) and for exogamous couples as in (5):

(4) *Percentage of endogamous couples = (Nmb of endogamous couples*$_{(FR+FR)}$*/Nmb of total couples) x 100%*

(5) *Percentage of exogamous couples = (Nmb of exogamous couples*$_{(FR+EN,\ FR+NO,\ FR+(EN+NO)}$*/Nmb of total couples) x 100%*

where FR, EN, and NO refer to each person's mother tongue.

Using the values calculated from equations (4) and (5), Figure 2 graphs the percentage of endogamous and exogamous couples in the regions. The data for 2001 and 2006 are for heterosexual couples only, while for 2011, 2016 and 2021, the results include hetero- and

homosexual couples. For Canada and New Brunswick, the two regions with a higher percentage of Francophones, we see a higher percentage of endogamous relationships compared to exogamous, whereas for Ontario and the Seven Provinces, we see a higher percentage of exogamous relationships compared to endogamous ones. While the trend for exogamous couples is not the same for all four regions, the trend for endogamous couples is consistently downward.

The drop in endogamous relationships could perhaps be related to the decrease of the proportion of the Francophone population observed in Section 3.1. Secondly, because we are only looking at unilingual Francophones and ignoring bilingual speakers, this could also be a reason for the downward trends in Figure 2.

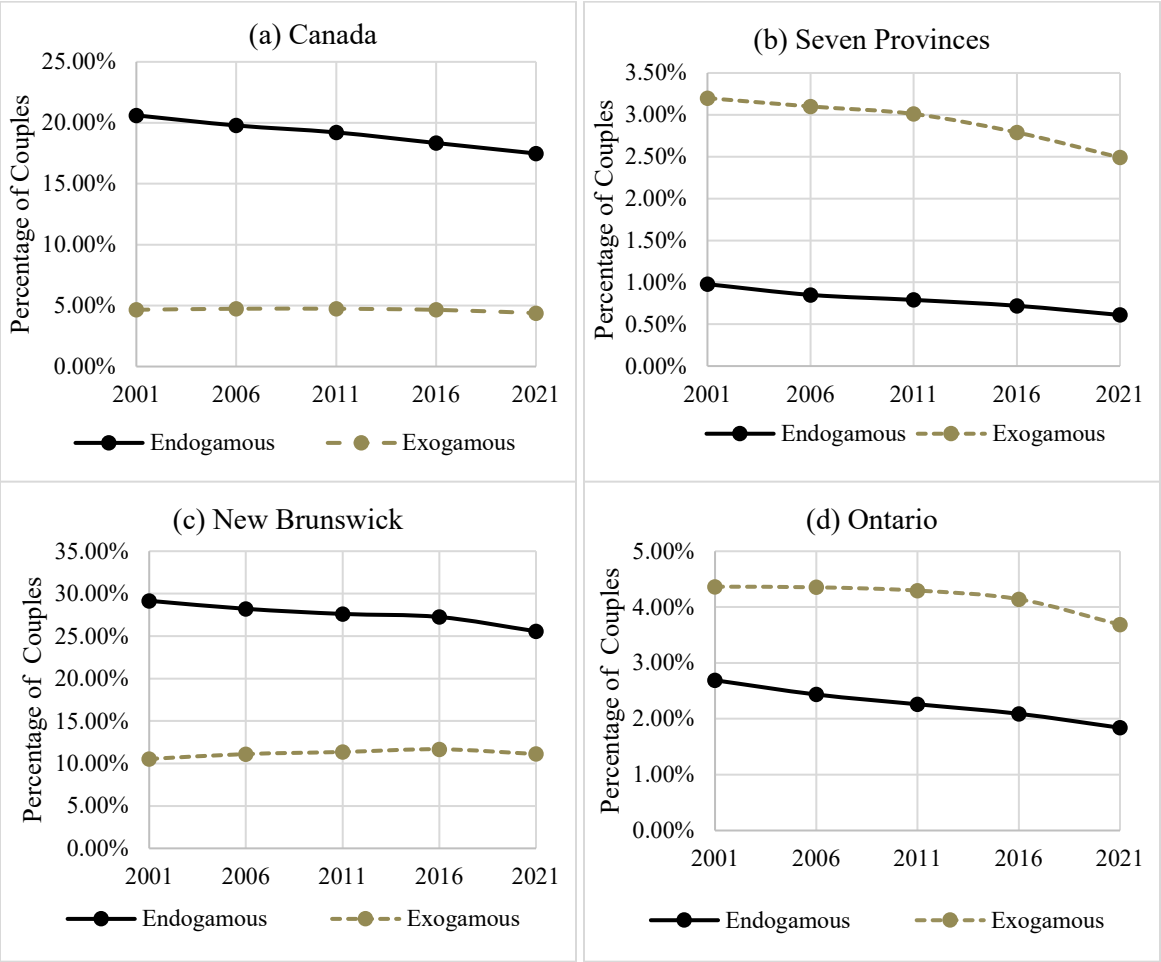


Figure 2 (a)-(d): Percentage of endogamous and exogamous couples for the 2001 to 2021 Canadian censuses for a) Canada, b) Seven Provinces, c) New Brunswick, d) Ontario

3.5 Language transmission

In Canada in 2021, over 98% of children born to an endogamous couple were Francophone, where only 0.9% of those children were unilingual Anglophone, and 0.6% were bilingual French/English. This gives a strong case that two Francophone parents will pass on French to their child/children. In this context, a decreasing percentage of endogamous couples observed in the previous section

is not a positive sign for French language transmission and, ultimately, the vitality of the Francophone population.

Table 3 offers a breakdown for the four regions of focus and presents the rate of unilingual and bilingual Francophone children born to endogamous couples. Data for the 2001 and 2006 censuses were not available on the Statistics Canada website, so only 2011-2021 data are presented. New Brunswick and Canada see 98.39% and 99.12% of children born to an endogamous couple identify as either unilingual or bilingual Francophones. Comparatively, we see lower rates in the Seven Provinces (between 82.47% and 85.65%) and in Ontario (between 91.45% and 92.32%). So, living in a region with a lower percentage of Francophones appears negatively related to the intergenerational transmission of French, even for families where both parents are Francophone.

Table 3: Percentage of unilingual and bilingual children (French and other), aged 0 to 17, born to endogamous couples for the Canadian census 2011 to 2021

Region	2011 (%)	2016 (%)	2021 (%)
Canada	99.12	99.09	98.99
Seven Provinces	84.55	85.65	82.47
New Brunswick	98.39	97.81	98.01
Ontario	92.32	92.28	91.45

We also note that for the four regions in our study, there is never a situation where 100% of children born to an endogamous (Francophone) couple have French as their mother tongue. We interpret this as a loss of Francophones which is not dependent on the fertility rate of the Francophone population.

Table 4 shows a table equivalent to Table 3 but for children born to exogamous couples. It is noticeable that the percentages across all regions are much lower in Table 4 than Table 3, indicating that the transmission of French for a child born to an exogamous couple is much lower than one born to an endogamous couple. The highest rate we see is 55.24% in 2021 for Canada. That means that only about half the children born to exogamous families in Canada learn French at home, and the Francophone population loses about half its speakers on the national scale. In the Seven Provinces, the rate of Francophone children is even lower at about a quarter (up to 27.36% in 2021), meaning only one in four children born to an exogamous relationship learn French as their mother tongue in Anglophone provinces. It is unclear why the percentages of French-speaking children increased for all four regions in Table 4 from 2011 to 2021. Our findings follow Mougeon (2014), who also observed an increase in the transmission of French from 1991 to 2001 and linked it to parental education and institutional support that contribute to a favourable view of status and socio-economic value of French (p. 220).

Table 4: Percentage of unilingual and bilingual children (French and other), aged 0 to 17, born to exogamous couples for the Canadian census 2011 to 2021

Region	2011 (%)	2016 (%)	2021 (%)
Canada	47.84	50.81	55.24
Seven Provinces	21.17	22.15	27.36
New Brunswick	40.71	43.21	45.31
Ontario	32.09	33.53	37.42

As we see, linguistic endogamy and exogamy play a large role in the replacement of the next generation of French speakers. With a decreasing percentage of endogamous couples in the four studied regions, there is a negative consequence for the transmission of French to the next generation. Independent of fertility rate, the mother tongue of two parents can greatly influence the mother tongue of their child.

4. Discussion

This study focused on determining language trends of the minority Francophone population in 21st century Canada using census data from 2001 to 2021. We examined the mother tongue, home language, and language shift trends in Canada, New Brunswick, Ontario and the Seven Provinces (the remaining Anglophone provinces).

We started by looking at 2001 and 2021 data for mother tongue and home language and calculating language shift from those values. With respect to our first research question that inquired on the demographic changes in the minority Francophone population, we found that the number of Francophones increased in Canada, Ontario, and the Seven Provinces, but decreased in New Brunswick, while the percentage decreased for all four regions, with a more noticeable change in Ontario and the Seven Provinces which have a lower concentration of Francophones. The results were similar for French users with an increase in the number of French users in Canada and the Seven Provinces and a decrease in the number of French users in Ontario and New Brunswick. Again, the percentage of French users decreased across the four regions, most notably in Ontario and the Seven Provinces. Regarding our second research question that queried on language shift trends, we observed that from 2001 to 2021, the language shift increased for the studied regions except for Canada, which can likely be attributed to Quebec having a large Francophone population and French as the official provincial language. We saw a more marked language shift in the regions with a lower concentration of Francophones and French users. Referencing our third research question, these trends seem to indicate that in regions with a higher concentration of Francophones, we saw a stronger linguistic vitality, whereas in regions with a lower concentration of Francophones, we saw a weaker linguistic vitality. Finally, we learned that in all four regions, the percentage of endogamic couples is steadily decreasing, and that the rate of exogamic couples is also lowering. This is notable because the rate of Francophone unilingual or bilingual children born to an endogamous couple in the four regions never reaches 100% and the rate of language transmission in exogamous households does not rise above 55.24%. So, to answer our fourth research question, exogamy could pose negative implications on intergenerational language transmission compared to endogamy and consequently governs the linguistic vitality of a region.

We saw consistent trends across all the regions of interest where the percentage of people with French as a mother tongue and French users decreased from 2001 to 2021. These results follow the trends observed by Castonguay (1993, 1997, 2001, 2005), DeVries (1994), and Langlois (2000). This is indicative of a decline in the health of the linguistic vitality of the minority Francophone population in the provinces outside of Quebec. Furthermore, living in an area with a lower concentration of Francophones shapes the minority community's way of living by creating a perpetual cycle of losing French speakers and not having enough of a community to support those still living in French. This is consistent with Castonguay's (2001) findings for census years 1971, 1981, 1991, and 1996 where New Brunswick and regions in Ontario with a higher Francophone density saw a lower language shift. People living in these localities also may have a

higher utility for knowing French compared to regions where the Francophone population is less dense, so Anglophones and Allophones may be more inclined to learn French because it is considered an asset.

A more detailed analysis should compare both the percentage and the numbers of endogamous and exogamous couples and include bilingual and multilingual Francophone parents for a study of the rate of language transmission.

For future work, it would be interesting to consider the language used at work (available for census years 2001, 2006, 2016, and 2021) and the language of instruction (data only collected in the 2021 census) to provide further insight on the language trends of the minority Francophone community in the 21st century. A larger study could also investigate how interprovincial migration and immigrants contributes to the changing values of mother tongue and home language, as well as looking at what official language Allophones may be shifting to (cf. Castonguay 2021).

From a demolinguistic perspective, the situation for Francophones in areas of less concentration is certainly precarious with declining rates of Francophones, French users, and intergenerational transmission of French.

References

Bergeron, Christian & Blanchet, Philippe & Lebon-Eyquem, Mylène. 2022. Étude exploratoire de l'insécurité linguistique et de la glottophobie chez des étudiants universitaires de l'Ontario. *Minorités linguistiques et société* 19. 3-25.

Bernard, Roger. 1997. Les contradictions fondamentales de l'école minoritaire. *Revue des sciences de l'éducation* 23. 509-526.

Bouchard-Coulombe, Camille. 2011. La transmission de la langue maternelle aux enfants : le cas des couples linguistiquement exogames au Québec. *Cahiers québécois de démographie* 40. 87-111.

Cartwright, Don. 1988. Language policy and internal geopolitics: The Canadian situation. In Williams, Colin H. (ed.), *Language in geographic context*, 238-266. Clevedon: Multilingual Matters.

Castonguay, Charles. 1993. Mesure de l'assimilation linguistique au moyen des recensements. *Recherches sociographiques* 1. 45-68.

Castonguay, Charles. 1997. Évolution de l'assimilation linguistique au Québec et au Canada entre 1971 et 1991. *Recherches sociographiques* 3. 469-490.

Castonguay, Charles. 2001. Démographie comparée des minorités de langue officielle. *Francophonies d'Amérique* 12. 25-36.

Castonguay, Charles. 2005. Vitalité du français et concentration des Francophones: un bilan 1971-2001. *Francophonies d'Amérique* 20. 15-24.

Castonguay, Charles. 2021. *Le fiasco de la politique linguistique canadienne.* Montreal : Les Éditions du Renouveau Québécois.

Corbeil, Jean-Pierre. 2005. L'exogamie et la vitalité ethnolinguistique des communautés Francophones en situation minoritaire : vécu langagier et trajectoires linguistiques. *Francophonies d'Amérique* 20. 37-49.

De Vries, John. 1994. Canada's official language communities: An overview of the current demolinguistic situation. *International Journal of the Sociology of Language* 105-106. 37-68.

Espenshade, Thomas J. & Guzman, Juan Carlos & Westoff, Charles F. 2003. The surprising global variation in replacement fertility. *Population Research and Policy Review* 22. 575-583.

Giles, Howard & Bourhis, Richard Y. & Taylor, Donald M. 1977. Towards a theory of language in ethnic group relations. In Giles, Howard (ed.), *Language, ethnicity and intergroup relations*, 307-348. London: Academic Press.

Hambye, Phillipe. 2019. La minorisation linguistique, entre discrimination et domination symbolique. Différences et enjeux de deux lectures des inégalités. *Minorités linguistiques et société / Linguistic Minorities and Society* 12. 15-30.

Jedwab, Jack. 2002. *Immigration and the vitality of Canada's official language communities: Policy, demography and identity*. Ottawa: Office of the Commissioner of Official Languages.

Landry, Rodrigue & Allard, Réal. 1994. The Acadians of New Brunswick: Demolinguistic realities and the vitality of the French language. *International Journal of the Sociology of Language* 105-106. 181-215

Langlois, André. 2000. Analyse de l'évolution démolinguistique de la population Francophone hors Québec, 1971-1996. *Recherches sociographiques* 2. 211-238.

Marquis, Marie-Hélène & Cormier, Marianne & Bigras, Nathalie. 2022. Liens entre le contexte sociolinguistique et les comportements langagiers de nouveaux parents : le cas du Sud-Est du Nouveau-Brunswick. *Éducation et francophonie* 50. 1-22.

Mougeon, Raymond. 2004. Perspective sociolinguistique sur le français en Ontario. In Coveney, Aidan, Hintze, Marie-Anne & Sanders, Carol (eds.), *Variation et francophonie*, 155-190. Paris : L'Harmattan.

Mougeon, Raymond. 2014. Maintien et évolution du français dans les provinces du Canada Anglophone. In Mufwene, Salikoko. S. & Vigouroux, Cécile (eds.), *Colonisation, globalisation et vitalité du français*, 211-276. Paris: Odile Jacob.

Mougeon, Raymond & Beniak, Édouard. 1994. Bilingualism, language shift, and institutional support for French: The case of the Franco-Ontarians. *International Journal of the Sociology of Language* 105-106. 99-126.

Provencher, Claudine & Galbraith, Nora. 2024. *Fertility in Canada, 1921 to 2022*. (https://www150.statcan.gc.ca/n1/pub/91f0015m/91f0015m2024001-eng.htm) (Accessed April 10, 2025).

Remysen, Wim. 2003. L'insécurité linguistique des Francophones ontariens et néo-brunswickois. Contribution à l'étude de la francophonie canadienne. In Langlois, Simon & Létourneau, Jocelyn (eds.), *Aspects de la nouvelle francophonie canadienne*, 95-116. Laval: Presses de l'Université Laval.

Statistics Canada. 2021-2022. *Dictionary, Census of Population, 2021. Appendix 2.1: Census of population and national household survey questionnaire content and derived variables since confederation.* (https://www12.statcan.gc.ca/census-recensement/2021/ref/dict/app/index-eng.cfm?ID=a2_1) (Accessed November 8, 2024).

Sheppard, Claude Armand. 1971. *The Law of Languages in Canada*. Ottawa: Information Canada.

Vézina, Mireille & Houle, René. 2014. La transmission de la langue française au sein des familles exogames et endogames Francophones au Canada. *Cahiers québécois de démographie* 43. 399-438.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:1.54-61 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Sauce Brune, Google Translate and ChatGPT; translating Quebecois swear words

Liza Kopylova
University of Waterloo

Abstract: This study examines the translation and retranslation processes of Google Translate and ChatGPT, specifically in relation to the play *Sauce Brune* by Quebecois writer Simon Boudreault. The play features Quebecois *sacres*, which are religious swear words, that pose a significant challenge to translate into English. To achieve this, a three-page monologue was put through Google Translate twice, and three times through ChatGPT. My primary goal was to compare the effectiveness of the two machine translation tools in translating Quebecois *sacres*. Secondly, I wanted to see if the back-translations produce Quebecois *sacres* similar to those used in the original monologue.

This paper presents two sentences from the original monologue and the six translated or back-translated versions including: the original, a Google Translate translation into English, a Google Translate retranslation into French, a ChatGPT translation into English, two versions of ChatGPT retranslation into French – the first closer to the French spoken in France, and the second closer to *Quebecois* slang. The study demonstrates that Google Translate and ChatGPT provide very different translations, and the *sacres* vastly differ in each respective version.

Keywords: translation, back-translation, Google Translate, ChatGPT, swear words

Languages: English, French

1. Introduction

1.1 Outline of the paper

THIS PAPER HAS THREE MAJOR OBJECTIVES. First, the priority is to examine the translation into English of French-language Quebecois swear words, also known as *sacres*. These swear words, unique to Quebec, are variations on the names of Catholic objects (Légaré & Bougaieff 1984: 2). They pose significant difficulty in translation studies due to their socio-cultural significance. Furthermore, this paper evaluates the effectiveness of automatic translation tools such as Google Translate and ChatGPT in the English-language translation of these swear words in the Quebec play *Sauce Brune* by Simon Boudreault. Finally, the paper explores back-translation, which is the retranslation of an already translated sentence into the original language, to see if these automatic translation tools will use the original *sacre* or substitute it with another French word.

1.2 Previous research

In the late-20th century, translators began using automated methods to expedite the translation of literary works, including novels and short stories. Today, researchers are examining the use of computer-assisted literary translation techniques, which include automatic translation tools, like Google Translate, and neural machine translation platforms that use machine learning, like ChatGPT. The majority of research on machine translation (Patil & Davies 2014: 7392) and artificial intelligence (Nadkarni et al. 2011: 544) concentrates on medical texts. It has been found that neural machine translation struggles to interpret handwritten medical notes (ibid.). Additionally, hospital patients who do not speak the official language may find that automated translations of medical brochures vary in quality depending on the target language (Patil & Davis 2014: 7392). Artificial intelligence tools, including ChatGPT, have some limitations, such as the inability to understand abbreviations, truncated words, or spoken language rules (Nadkarni et al. 2011: 544). As a relatively new tool, ChatGPT, which became accessible to the public in 2022, has received less academic attention and research (Breen 2023). The studies that do exist indicate that the accuracy of the translation varies from one language to another.

Moreover, artificial intelligence lacks awareness of the socio-cultural subtleties that shape a language or its use in literature. It sometimes fails to find an equivalent when translating idiomatic or sociolectal expressions and phrases. According to Legaré & Bougaieff (1984: 2), the Quebec expletive swearword, or *sacre*, holds sway over all social strata. It derives from the Catholic religion brought by French colonization, yet its usage typically takes on different, modified forms, rather than directly adopting the original sacrosanct terms. Quebecers do not always use the same intensity in their derivatives of sacred words. Some of these derivatives and interjections, such as *nom de Dieu* and *bon Dieu,*[1] are not considered blasphemous today, but function rather as invocations to the sacred (p. 9). They can be used in different ways, such as interjections, appositions, emphasis/threat, or adverbs (1984: 20).

Translating Quebecois *sacres* poses a significant hurdle, though. To our knowledge, there are no studies on the translation of swear words or of Quebecois *sacres* using tools of machine translation or neural machine translation. There is also a lack of research on the translation or usage of these expressions in Canadian theatre, specially using machine translation or neural machine translation. As Louise Ladouceur points out in her book *Making the scene: Translating theatre from one official language to another in Canada*, when faced with Quebec swear words, theatre translators employ various strategies to cater to both readers and viewers. They can also prioritize emotion over meaning, as Ladouceur (2005: 61) suggests, to evoke a similar impact as the original message.

1.3 Outline of the study

To test the limits of machine translation and neural machine translation, I focused on a three-page monologue of the Quebec play *Sauce Brune*, which is hyperbolic in its use of *sacres*. I used machine translation and neural machine translation tools to translate this monologue from French into English: once with Google Translate, and once with ChatGPT. I then used the same tools to retranslate the translated monologue to evaluate the retranslations against the original monologue. Two different translations were generated using ChatGPT, the first using the instruction "Translate

[1] These roughly translate to 'in the name of God' and 'good God.'

into French," and the second "Translate into Quebec French."[2] This paper presents two sentences and all their respective versions.

2. The play *Sauce Brune*

2.1 Summary of the play

Published in 2010, *Sauce Brune* by Simon Boudreault features four cafeteria workers at a high school. They talk about what they are going to cook for the students, while insulting each other, especially picking on Martine, the youngest and most naïve of them all. The play begins with a daily menu announcement, increasing the pressure on the characters to create new dishes while catering to each student's dietary needs. The women constantly compete, each trying to convince the others that her ideas are superior. In an article from the daily newspaper *La Press*, Blais (2009) writes that the play is an ode to Quebec writer Michel Tremblay, who often explores the hardships faced by working-class women, using profanity to convey their inner turmoil. Boudreault takes this concept further, enhancing the vulgarity of his characters by using Quebec slang and swear words to their fullest potential.

2.2 Examples used in *Sauce Brune* and definitions

For the research presented here, I chose a three-page monologue portrayed by the character Martine in the play *Sauce Brune*. Martine, who is the youngest of the team of canteen girls, is always a victim of the other women. They blame her for every mistake and discourage her ideas for the menus. After suffering the insults of the others, Martine finally explodes and expresses all her previously suppressed emotions and thoughts in the monologue. She swears in reference to the other characters. The Quebecois swear words she uses highlight the intensity of Martine's emotions and her feelings towards her colleagues. Martine is not insulting religion or God; she is insulting others and intensifying her sentences.

Three Quebecois *sacres* are used in this monologue: they are variations of *câlisse* (*câlissemment*), of *ostie* (*estie*, *stie*), and of *crisse* (*crisser*, *crissement*). The word *câlisse* 'calice' refers to the sacred vessel in Christian religions (Usito). Meanwhile, the word *ostie* is a variation of *hostie*, which according to the Quebec French dictionary Usito is "[i]n the Catholic liturgy, a small thin slice of unleavened bread that the priest consecrates during Mass" (translated into English by the author from Usito).[3] Finally, the words *crisse* and *crissement* are variants of *Christ*. These variations do not have the literal meaning that would be used in a church setting; they are rather emotional additions to the character's sentences.

2.3 Deforming tendencies of literary translation

Since it is plausible that Google and ChatGPT's translations may fall short, especially when it comes to *sacres* and vernacular components, which represent the pinnacle of linguistic complexity, I examined the outcomes using the framework of "thirteen deforming tendencies" of literary translation proposed by Antoine Berman (1999: 49). These deforming tendencies include:

[2] The commands in ChatGPT were done in French. "Translate into French" was "Traduit ce texte vers le français" and "Translate into Quebec French" was "Traduit vers le français québécois."

[3] The original definition in French is: "Dans la liturgie catholique, petite rondelle mince de pain azyme que le prêtre consacre pendant la messe."

1. Rationalisation, which is the decomposition of sentences and passages into their constituent parts to understand the logical sequence of the text.
2. Clarification introduces a sense of definiteness, imposing structure on what was once undefined (p. 54-55).
3. Lengthening is where the translated sentence is longer than the original.
4. Ennoblement and vulgarisation. The first prioritizes making the text more "elegant" and "poetic" than that of the original. In contrast, vulgarization refers to the use of everyday vernacular in written form (p. 57–58).
5. Qualitative impoverishment occurs when words, phrases, and other linguistic elements are replaced by others that lack the same depth of meaning, sound, or significance (p. 58).
6. Quantitative impoverishment is the decrease in lexical diversity and the increase in the number of signifiers (p. 59).
7. Homogenization is the unification of the original heterogeneous text, resulting in a combination of all the previous tendencies (p. 60).
8. Destruction of rhythms, including punctuation.
9. Destruction of underlying signifying structures, hidden meanings connected from afar, forming different facets of the text's significance.
10. Destruction of systematisms extends beyond mere signs to encompass grammatical structures such as tense usage, subordinate clause varieties, and so on.
11. Destruction or exoticization of vernacular language networks, leading to the shift from vernacular to cultivated language. In discussing the destruction or exoticization of vernacular language networks, Berman argues that one vernacular cannot translate another, and that translation ridicules the original version (p. 64).
12. Demolition of locutions, distorting the original meaning of proverbs and leading to widespread confusion.
13. Erasure of multilingualism, which translation reduces to a single language (p. 66).

The translations of the play *Sauce Brune* do not allow us to provide an example of every single one of these deforming tendencies; however, in the analysis portion of this study, I identify which "deforming tendencies" best characterize the translations and back-translations provided by Google and ChatGPT.

3. Analysis

In the analysis portion of this paper, two sentences are presented with all five of their respective translations from ChatGPT and Google Translate to show which deforming tendency these tools adhere to. These two sentences presented here have been chosen from the monologue as they are longer and demonstrate the most variety of swear words.

3.1 Analysis of the first sentence translated into English by Google and ChatGPT

Table 1 shows the original sentence with the two English translations. In the first translation offered by Google, the variants of *Christ* are translated into words that are English-language synonyms of the French word *cris*, (and not *Christ*), such as *scream*, *squeal*, and *screech*. This phenomenon is known as false friends which "are generally defined as pairs of words that look or sound phonologically and/or orthographically similar in two languages but differ significantly in meaning." (Al-Athwary 2021: 368) Despite their similarity, as they are verbs and they signify a

harsh sound, these words do not carry the same meaning in French and English. False friends recur often throughout the Google Translate version and lead to the destruction of underlying structures. As mentioned in Section 2.3, this deforming tendency destroys the underlying network of similar words. In every instance, the absence of profanity in English dulls the exaggerated tone of Martine's monologue.

Table 1: First sentence translated by Google and ChatGPT

Original	**Google Translate**	**ChatGPT**
Faque la pauvre conne de fille a sait pus quoi faire, **estie**, a reste dehors à s'faire fouetter par l'**ostie** d'pluie, pis a sait **crissement** pus c'est quoi qu'y a de pire entre garder ses **crisses** de barniques ou les **crisser** à l'autre boutte de **l'ostie de planète, crisse**	Because the poor stupid girl doesn't know what to do, she stays outside getting whipped by the rain shower, and she knows how to **scream** and what's the worst between keeping her **screams** barniques or **squeal** them at the other end of the **ostia** of the planet, **screech**	So the poor damn girl doesn't know what to do anymore, **damn** it, she stays outside getting whipped by the **damn** rain, and she **damn** well doesn't know what's worse, whether to keep her **damn** sunglasses or **throw them** to the other **damn** end of the planet, **damn it**

In the overall English translation given by ChatGPT, the neural machine translation tool translated all variations of the three *sacres* as *damn*. This occurred 162 times throughout the three-page monologue. Of the seven variations in the first example alone, six out of seven were translated to *damn*. Despite its religious reference, *damn* does not hold the same weight in English: "*Damn you!* A powerfully taboo body-related word substitutes for the weakly taboo religious one so that the curse can retain its impact." (Mohr 2013: 44)[4] ChatGPT's excessive use of the word *damn* leads to a quantitative impoverishment (Berman's deforming tendency that distorts the lexical richness of *sacres*) in the original monologue. The repetition of this word also has the effect of homogenizing the final text. Already quite weak, this word can make the final text monotonous.

3.2 Analysis of the back-translation of the first sentence

Table 2 demonstrates the back-translations in Google Translate and ChatGPT. The Google Translate back-translation is the shortest, with only 28 words. This translation replaces the false friends with their French counterparts: *fouetter*, *crier*, *cri*; none of these words are *sacres*. The destruction of vernacular networks, evident in the English translation (Table 1), is further confirmed by the back-translation, which only retains *ostia* among the seven *sacres*.

[4] In English, blasphemous swear words have existed since at least the Middle Ages, but they were considered more profane at the time: "In the *bad* sense, an oath means blasphemous or vain swearing, words or phrases that take God's name in vain, mention his body parts, or otherwise detract from his honor." (Mohr 2013: 7-8)

Table 2: Back-translations of the first example

Google Translate	ChatGPT 1	ChatGPT 2
Parce que la pauvre idiote ne sait pas quoi faire, elle reste dehors à se faire **fouetter** par la pluie, et elle sait **crier** et quel est le pire entre garder **ses cris** dans des barniques ou les **pousser** à l'autre bout de **l'ostia du planète**, **cri**	Alors la pauvre putain de fille ne sait plus quoi foutre, **bon sang,** elle reste dehors se faisant fouetter par la **foutue** pluie, et elle ne sait **foutrement** pas ce qui est pire, garder ses **foutues** lunettes de soleil ou les **foutre** à l'autre **foutu bout de la planète, bon sang**	Donc la pauvre maudite fille sait plus quoi faire, **crisse**, elle reste dehors à s'faire battre par la **maudite** pluie, pis elle sait ben pas c'est quoi l'pire, garder ses **maudites** lunettes de soleil ou les **lancer** à l'autre **maudit bout de la planète, crisse**

In contrast, the two ChatGPT back-translations do not align with the original English translation. Since the English one was overly redundant and excessive with the word *damn*, both back-translations provide a diverse selection of expletives. In its initial form (under the column "ChatGPT1"), this text contains French profanities. Subsequently (under the column "ChatGPT2"), a range of Quebecois *sacres* are introduced. The deforming tendency seen here is vulgarisation. As the overuse of *damn* rendered the text monotonous, the variety seen in these two versions renders the text more vulgar and akin to the intensity of the original Quebecois play. Of the seven *sacres*, all were translated in the first ChatGPT back-translation: *bon sang*, *foutue, foutrement, foutues, foutre, foutu bout de la planète,* and *bon sang*. Although *bon sang* is repeated twice, the other five are all variations of *foutre*; however, it is not redundant in nature. Moreover, in the second back-translation, six of the seven sacres were translated: *crisse, maudite, maudites, lancer, maudit bout de la planète,* and *crisse.* What is important to note, in regard to ChatGPT, is that when inputting the simple command "Translate back to French," it provided a French dialect. This also includes implementing swear words that are native to the French dialect. So, of the previously mentioned examples, such as *foutre*, this was a French swear word. Therefore, I decided to conduct a second back translation in ChatGPT, in which the command specified to "Translate to Quebecois French." When doing so, the provided text had Quebecois *sacres*. Lastly, only *lancer* is neither a swear word nor a *sacre*; however, the other five preserve the vulgarity of the sentence.

3.3 Analysis of the second sentence translated by Google Translate and ChatGPT

Table 3: Second sentence translated by Google Translate and ChatGPT

Original	Google Translate	ChatGPT
Rien a connaît ça, a n'a eu **crissement** de rien, a n'en manque pas de d'ça, pis à l'aime mieux ça parce que des fois**, estie,** des **osties** d'fois d'**estie** de **câlisse** est **crissement** ben avec c'te **crisse** d'écœurant d'sale-là	Nothing knows that, has never missed anything, has no shortage of that, and likes it better because sometimes, **estie, osties** of times of **estie** of **calisse** is **screeching** well with that disgusting **screeching** sound	Nothing knows it, **damn** it, she **damn** well has nothing, she doesn't damn well miss anything, and she **damn** well prefers it because sometimes, **damn** it, **damn** times **damn** well are **damn** well with that **damn** disgusting damn[5] pig

[5] This particular damn refers to the originals *d'sale-là* which is not a sacre. Therefore, it was not analyzed.

This second sentence also has seven *sacres* (Table 3). This one also implements *câlisse* which is the rarest in the monologue. In the Google Translate version, there are six translations: *estie, osties, estie, calisse*, and *screeching* twice. The word *screeching* is another example of a false friend. However, the other words were imported directly from French. This reaffirms the destruction of vernacular networks as the sentence in English is now illogical and not profane.

In The ChatGPT English translation, all seven *sacres* were translated into *damn*. To reiterate, the English ChatGPT version is an example of quantitative impoverishment.

3.4 Analysis of the back-translations of the second sentence

Table 4 offers the three back-translations of the second example. In the Google Translate example, there are five instances of the use of *sacres*: *estie, osties, estie, calisse* and *cri*. *Cri* is the translation of the false friend *screeching*, therefore neither a *sacre* nor a swear word.

In the two ChatGPT versions, much of the sentence was condensed. The first version has just three swear words: *foutrement*, twice, and *bon sang*, whereas the second has three instances of *crisse*. *Foutrement* and *bon sang* are both French vernaculars whereas *crisse* is a *sacre*. Despite the repetition, like *damn* in the English version, the fact that there are only three swear words in each respective sentence renders it not as redundant as the seven uses of *damn* in the English version.

Table 4: Back-translations of the second example

Google Translate	**ChatGPT 1**	**ChatGPT 2**
Rien ne sait ça, n'a jamais rien raté, n'en manque pas, et aime ça mieux parce que parfois**, estie, osties** des temps d'**estie** de **calisse** hurle bien avec ce **cri** dégoutant	Rien le sait, bon sang, elle n'a **foutrement** rien, elle ne foutrement manque de rien, et elle foutrement préfère ça parce que parfois, **bon sang**, parfois **foutrement** bien fout	Rien sait, **crisse,** elle a rien, elle manque de rien, et elle préfère ça parce que des fois, **crisse,** des fois **crisse** ça va ben

4. Conclusion

Based on the two examples analyzed, only four deforming tendencies were identified. The destruction of vernacular networks was a prominent feature of the English Google Translate sentences. The machine translation relied on false friends and was unable to find the English equivalents of the *sacres*. The English version of ChatGPT had a high level of quantitative impoverishment due to overuse of the word *damn*. For a native English speaker, this translation is not intense because of this overuse. This translation is also an example of homogenization, as it is monotonous. Furthermore, the two versions of ChatGPT demonstrate vulgarization, despite being derived from a text with 162 instances of the same word. Interestingly, both translations incorporate a diverse range of French vernaculars. As discussed, with ChatGPT it is crucial to specify the prompt when asking it to translate a text. With the simple command "Translate into French," the artificial intelligence defaulted to the French dialect by utilizing French swear words. By specifying the prompt to "Translate to Quebecois French" in the second back-translation of the text, ChatGPT used Quebecois *sacres*. This demonstrates that ChatGPT is capable of using multiple dialects; however, its default is the French from France and not of the French from Quebec.

This study demonstrates that ChatGPT outperforms Google when it comes to translating Quebecois profanities in a dramatic text. However, none of these tools can match the frequency and intensity of the swear words of Martine's monologue in *Sauce Brune*. Additionally, it has been established that ChatGPT requires explicit instructions to prioritize Quebec colloquialisms during reverse translation, as default settings favour French slang. To delve deeper into this analysis, we could examine additional profanities, including sacred oaths, expletives, and Quebec vernacular. An extended extract from the play, such as an entire act in Boudreault's play, might provide valuable insights on other deforming translation tendencies. Such an exercise would shed light on other potential challenges and opportunities associated with adapting artificial intelligence and machine translation tools to regional dialects.

References

Al-Athwary, Anwar A.H. 2021. False friends and lexical borrowing: A linguistic analysis of false friends between English and Arabic. *Journal of Language and Linguistic Studies* 17. 368-383.

Berman, Antoine. 1999. *La traduction et la lettre, ou, l'auberge du lointain.* Paris: Éditions du Seuil.

Blais, Marie-Christine. 2009. *Sauce brune : « crissement » intéressant. La Presse*. https://www.lapresse.ca/arts/spectacles-et-theatre/theatre/200903/25/01-839949-sauce-brune-crissement-interessant.php

Boudreault, Simon. 2010. *Sauce Brune.* Montréal: Dramaturges Éditeurs.

Bowker, Lynne. 2021. *Vous traduisez pour le Canada?* Version 2. Ottawa: University of Ottawa.

Breen, Benjamin. 2023. *Translating Latin Demonology Manual with GPT-4 and Claude. Res Obscura.* https://resobscura.substack.com/p/translating-latin-demonology-manuals

"Calice." *Usito.* Université de Sherbrooke. https://usito.usherbrooke.ca/d%C3%A9finitions/calice

"Hostie." *Usito.* Université de Sherbrooke. https://usito.usherbrooke.ca/d%C3%A9finitions/hostie

Kraif, Olivier & Roux, Pascale. 2022. Comparaison d'un texte original et de ses rétrotraductions : que disent les mesures textométriques? *Meta* 67. 190-216.

Ladouceur, Louise. 2005. *Making the scene. La traduction du théâtre d'une langue officielle à l'autre au Canada.* Quebec: Éditions Nota Bene.

Legaré, Clément & Bougaieff, André. 1984. *L'empire du sacre québécois étude sémiolinguistique d'un intensif populaire.* Quebec: Presses de l'Université du Québec.

Mohr, Melissa. 2013. *Holy sh*t: A brief history of swearing.* Oxford: Oxford University Press.

Nadkarni, Prakash M. & Ohno-Machado, Lucila & Chapman, Wendy W. 2011. Natural language processing: An introduction. *Journal of the American Medical Informatics Association: JAMIA* 18(5). 544-551.

Patil, Sumant & Davies, Patrick. 2014. Use of Google Translate in medical communication: Evaluation of accuracy. *British Medical Journal* 349(1). eArticle g7392.

This article was first published at https://www.lacussquare.org

LACUS FORUM

Journal of the Linguistic Association of Canada and the United States

VOLUME 50 NUMBER 2, 2025

CONTENTS

6. Between predictability and randomness: Seeking artistic inspiration from AI generative models — 1
Olga Vechtomova

7. The rhetoric of the *Ukrainization of Poland*: Mechanisms and effects — 13
Grażyna Drzazga & Magda Stroińska

8. On the success and limitations of a quantitative study of geographic and dialect distance — 24
Sheila Embleton & Eric S. Wheeler

9. Prepositions in Mixtec? The case of gradient word classes in Tu'un Javi — 35
DeAndré A. Espree-Conaway

10. Morphological complexities and socio-linguistic barriers against the adoption of feminatives in contemporary Polish discourse from the Gen Z perspective — 51
Artur Urbaniak

LACUS Forum **50:2.1-12 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Between predictability and randomness: Seeking artistic inspiration from AI generative models

Olga Vechtomova
University of Waterloo

Abstract: Artistic inspiration often emerges from language that is open to interpretation. This paper explores the use of AI-generated poetic lines as stimuli for creativity. Through analysis of two generative AI approaches – lines generated by Long Short-Term Memory Variational Autoencoders (LSTM-VAE) and complete poems by Large Language Models (LLMs) – I demonstrate that LSTM-VAE lines achieve their evocative impact through a combination of resonant imagery and productive indeterminacy. While LLMs produce technically accomplished poetry with conventional patterns, LSTM-VAE lines can engage the artist through semantic openness, unconventional combinations, and fragments that resist closure. Through the composition of an original poem, where narrative emerged organically through engagement with LSTM-VAE generated lines rather than following a predetermined structure, I demonstrate how these characteristics can serve as evocative starting points for authentic artistic expression.

Keywords: artificial intelligence (AI), creativity, generative AI models, artistic inspiration, Human-AI Collaboration, Large Language Models (LLMs)

1. Introduction

ARTISTIC INSPIRATION IS A COMPLEX PHENOMENON that often emerges from the interplay between predictable and unpredictable elements. In the context of the Seed phase of creativity, where initial ideas spark the creative process (Rubin 2023), artists have long employed techniques to disrupt conventional patterns and create new possibilities for expression. With the recent advances in artificial intelligence (AI) research, new opportunities arise for generating such creative stimuli through neural text generative models.

This paper investigates how AI-generated poetic lines can serve as sources of inspiration during the Seed phase, focusing particularly on their potential to balance structure and unpredictability. I examine two different approaches to the use of AI in poetry creation: individual lines generated by Long Short-Term Memory Variational Autoencoders (LSTM-VAE) that catalyze human creative engagement, leading to emergent narratives and original works, and complete poems autonomously generated by Large Language Models (LLMs). Through analysis of two poems – one that emerged through engagement with LSTM-VAE generated lines, and another by an LLM – I demonstrate how LSTM-VAE are particularly effective as sources of creative inspiration.

LSTM-VAE models, due to their design and training objectives, generate text that occupies a productive middle-ground between randomness and predictability. Their output combines semantic openness with unconventional linguistic patterns, enabling artists to discover and develop original poetic narratives[1] through genuine emotional engagement with the generated lines. In contrast, while LLMs produce technically accomplished poetry, their adherence to conventional forms tends to constrain rather than encourage creative interpretation. The contributions of this paper are threefold:

1. An examination of how AI can extend historical artistic techniques of creative disruption, establishing a framework for understanding a machine-generated text as a catalyst for artistic discovery.
2. A comparative analysis of LSTM-VAE and LLM approaches to poetry generation, demonstrating how their different architectures influence their potential for inspiring creativity.
3. Evidence for how specific qualities of AI-generated lines – particularly their balance of structure and unpredictability – can enable the emergence of authentic artistic narratives.

This investigation reveals how specific qualities of AI-generated lines – their linguistic patterns, semantic openness, and capacity for novel combinations – influence their potential as creative stimuli. The organic emergence of poetic narrative through engagement with LSTM-VAE lines demonstrates that productive linguistic innovation, rather than mere technical accomplishment, can inspire an artistic process. These findings have important implications for developing AI systems that can effectively support human creativity.

2. Inspiration and the creative process

The creative process in artistic endeavors, particularly in poetry and lyric writing, involves distinct phases that require different cognitive and artistic approaches. Rubin (2023) formulates three key stages:

Seed phase. This initial stage represents a state of creative receptivity where diverse stimuli can spark artistic inspiration. It involves heightened sensitivity to linguistic input (words, phrases, syntactic structures), openness to unexpected associations, and engagement with ambiguous elements that can trigger multiple interpretations.

Experimentation. In this intermediate phase, artists combine and reconfigure elements discovered during the Seed phase, engaging both generative and evaluative aspects of creative thinking. This phase represents a convergence of the initial inspiration and artistic vision, where raw elements take coherent form.

Crafting. The final stage involves applying technical skill and artistic judgment to refine the work. This requires both mastery of technique and sensitivity to the original creative impulse that emerged during the Seed phase.

While these stages are conceptually distinct, they can overlap in practice, with creative inspiration potentially emerging throughout the process. This study focuses specifically on the Seed phase and how AI-generated lines can serve as effective creative stimuli during this crucial initial stage. I argue that certain qualities of generated text – particularly *unpredictability* and

[1] In this paper, I use *narrative* in a broad sense, not merely as a traditional storyline with distinct characters and events, but as an emergent trajectory of meaning or emotional arc that arises from the arrangement of poetic lines.

semantic openness – can be especially valuable during the Seed phase, where disruption of habitual patterns often leads to a creative breakthrough.

The effectiveness of unpredictable or indeterminate elements during the Seed phase aligns with established artistic practices throughout history. As will be discussed in the next section, many 20th-century artists, poets, and writers deliberately employed techniques that introduced randomness and unpredictability into their creative process, particularly during the initial stages of artistic creation.

This historical context helps us understand why certain characteristics of AI-generated lines – such as syntactic estrangement, deliberate incompleteness, and unconventional semantic relationships – might make them particularly effective as seeds for artistic inspiration. Just as traditional techniques like the cut-up method created space for new artistic possibilities, AI-generated lines with the right balance of unpredictability and coherence can serve as valuable starting points for creative exploration.

3. Unpredictability in the creative process

Artists have long sought methods to break free from conventional language patterns that can restrict creative expression. The cut-up technique, originally proposed by the Dadaists in the early 20th century, was one such method. It consisted of literally using scissors to cut up a page of printed text into words and letting them fall in a random order. While the Dadaists intended this as a statement challenging the very notion of art, William Burroughs later popularized it as a literary technique. Burroughs (2012) argued that language is a "lock" that restricts creativity and confines us to predictable patterns. Musicians like David Bowie (Jones 2017) and Kurt Cobain (Cross 2001) adopted this technique in their songwriting. Bowie described using cut-ups to stimulate his imagination through random juxtaposition of concepts, stating: "What I've used it for, more than anything else, is igniting anything that might be in my imagination." (BBC 1975)

Syntactic defamiliarization – the deliberate disruption of expected syntactic patterns – emerged as another powerful technique for achieving creative unpredictability. Russian Formalist Viktor Shklovsky's (1965) concept of *ostranenie* – 'estrangement' or 'defamiliarization' – provides a theoretical framework for understanding how breaking conventional syntactic-semantic relationships can make language feel fresh and strange. When artists deliberately create syntactic *errors* or incomplete structures, they force readers to engage with a language in new ways, much like the cut-up technique. This syntactic estrangement prevents automatic processing of language and demands active meaning-making from the audience.

The intentional use of randomness and unpredictability extended beyond literature into visual art of the early to mid-20th century, with artists like Jean Arp, Georg Nees, Kurt Schwitters, and Vera Molnar incorporating random elements into their work. Georg Nees, considered to be the author of the first computer art, noted that randomness alone cannot be the source of art – there must be an aesthetic structure to narrow down the outcomes and produce discernible aesthetic information (Nees 1969). Molnar observed that computer programs with random generators help artists explore a broader range of possibilities in finding the right artistic form (Molnar 2019).

Kurt Schwitters (2021) experimented with word unpredictability in his Merz-poems through collage techniques, incorporating various elements from different sources. His deliberate linguistic disruptions aimed to heighten emotional impact, as he explained regarding his poem *An Anna Blume*: ""I love thou" is more expressive than "I love thee," because it's unusual. Had I written "I love thee" nobody would have noticed that I love Anna Blume." (Schwitters 2021: 48)

This tradition of intentional language disruption continued with the *Language poetry* movement of the 1970s, which posited that language dictates meaning rather than the opposite. Poets like Bob Perelman (1978) deliberately broke up language to engage readers in actively constructing meaning. More recently, novelist Tom Spanbauer's "burnt tongue" technique – deliberately misusing language to make it fresh – echoes these earlier approaches. As his student Chuck Palahniuk explains, it is "a way of saying something, but saying it wrong, twisting it to slow down the reader." (Palahniuk 2002) Eye-tracking studies validate this approach, showing that predictable words are indeed skipped more often by readers and elicit shorter eye fixations (Lowder et al. 2018). For example, readers spend less time on *cake* in *the baker rushed the wedding cake* than on *pies* in *the baker rushed the wedding pies*.

4. Predictability in art

Predictability in art has served both as a target of critique and as a deliberate artistic device. Artists have often employed predictable elements, triviality, and cliché as part of intentional artistic statements, using familiarity itself as a tool for engaging audience expectations.

The pop-art movement, which emerged in the 1950s and 1960s, is one of the best-known examples of an art movement using predictability and banality as a device, often with an intent to critique the superficiality of consumer culture. Artists such as Andy Warhol and Barbara Kruger often used recognizable imagery from advertisements, comics, and popular culture with intentionally cliché and banal statements, such as "I shop therefore I am." (Kruger 1990)

The conceptual art movement of the 1960s further explored predictability as an artistic device. In conceptual art, where the concept behind the work is prioritized over its aesthetic or material qualities, banality became a tool for challenging viewers' perceptions of art, language, and everyday life. Jenny Holzer's *Truisms* (Holzer 1978-1987) exemplifies this approach, using everyday language and declarative sentences to create intentionally trivial-sounding statements like "a little knowledge goes a long way" or "action causes more trouble than thought."

While pop-art and conceptual art deliberately employed predictability and banality as artistic devices, in literature, predictability is often identified as inauthentic and a form of artistic short-cut. This perspective is captured in Tom Spanbauer's concept of "received text." As Chuck Palahniuk (2011) explains:

> To Tom [Spanbauer], "wringing your hands" would be called "Received Text." Like a cliché, but more subtle. The phrase might not be as bad as "warm as a summer's day" or "pretty as a picture," but the phrase was still a short-cut, or pathway well-worn by a lot of writers and easy to feed into a story. (Palahniuk 2011)

5. Complexity and aesthetic perception

The relationship between aesthetic pleasure and complexity, known as the inverted U-shaped or *Wundt* effect, suggests that stimuli of intermediate complexity are most pleasing. This effect, first explored by Berlyne (1975), has been observed in various domains such as music (Gold et al. 2019) and product design (Althuizen 2021).

Complexity is typically expressed in features such as predictability, surprise, and uncertainty. Gold et al. (2019) offered a possible explanation for the inverted U-shaped effect, grounded in how the brain processes and learns from information. When language or other stimuli are completely

predictable, there is nothing new to learn. Conversely, when they are entirely random, no meaningful patterns can be discerned. It is in the middle-ground – where patterns are discernible but not obvious – that the brain engages most deeply, activating reward mechanisms that make the experience pleasurable. While there is evidence that the U-shaped effect exists across different aesthetic domains, there is also a strong indication that our preferences of aesthetic stimuli are highly individual (Gold 2019; Althuizen 2021).

Building on these insights, I hypothesize that short lines of text in the mid-region of the random-predictable spectrum are most likely to be useful as seeds for ideas and inspiration during the Seed phase of the creative process. In the next section, I introduce two generative AI architectures, LSTM-VAE and LLMs, and in the subsequent section I discuss why LSTM-VAE is more suitable as a source of evocative lines during the Seed phase compared to LLMs.

6. Neural text generative models

Researchers in Natural Language Processing have worked on text generative models for decades with varying success. The first major breakthrough took place around 2016 with the successful adoption and development of neural network architectures based on Recurrent Neural Networks (RNN), especially their more effective variant, Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber 1997). The second major breakthrough came in the early 2020s with the introduction of Large Language Models based on the Transformer architecture (Vaswani et al. 2017).

It is important to clarify in this context what a generative natural language model is. It is a model that is capable of producing original sequences with or without conditioning in the form of a text (e.g. prompt in an LLM) or any other signal (e.g. image, audio, video). The word *original* here is used without any value judgement describing the novelty or aesthetic merit of the text. It simply refers to a text that is not directly derived from the input, for example, replication of the input or any other hard-coded or rule-based response.

This section introduces two neural network architectures for text generation: Variational Autoencoders based on Long-Short Term Memory Networks (LSTM-VAE) and Transformer-based Large Language Models (LLMs).

6.1 LSTM-VAE: Generating linguistic variability

Before the invention of the Transformer-based Large Language Model (LLM), one of the most effective neural architectures for text generation was the Variational Autoencoder (VAE) (Kingma & Welling 2014). LSTM-VAE refers to a Variational Autoencoder that uses Long Short-Term Memory networks. This model is particularly well-suited for generating short, evocative sentences that balance randomness and predictability, making it valuable for creative applications. At its core, the LSTM-VAE consists of two main components:

Encoder: This part of the model reads an input sentence and compresses it into a condensed representation known as the *latent vector*. Think of this as capturing the essence of the sentence – its underlying semantic, stylistic, and syntactic features – in a numerical form. The domain in which latent vectors exist is called the *latent space.*

Decoder: This component takes the latent vector and attempts to reconstruct the original sentence from it. Essentially, it translates the compressed representation back into natural language.

During training, the model learns to encode sentences into the latent space and then decode

them back to their original form. The key innovation of the VAE is that it encourages the latent space to be smooth and continuous. This means that small changes in the latent vector result in small changes in the generated sentence, allowing for controlled variability.

By sampling different points from the latent space, we can generate new, original sentences that are similar but not identical to those in the training data. Because the latent space is structured, sentences generated from nearby points in the space will share certain characteristics, such as themes or stylistic elements.

After the LSTM-VAE is trained, we can generate new sentences by following these two steps:

Sampling: Randomly selecting a point in the latent space. This introduces an element of randomness, allowing for the creation of novel sentences.

Decoding: Feeding this point into the decoder to generate a sentence. The decoder transforms the latent vector back into a sequence of words, producing a coherent sentence.

Unlike later Large Language Models, LSTM-VAE is only able to generate short sentence-length texts. However, an important advantage of LSTM-VAE is that it does not require massive amounts of data to produce interesting results. Because it focuses on capturing the essential features of sentences, it can be trained on smaller, curated datasets. This allows for greater control over the style and content of the generated text.

Furthermore, LSTM-VAE can be extended to include additional information – a variation known as the Conditional Variational Autoencoder (CVAE). For example, we can condition the model on an external signal, such as an audio clip or a specific theme, to guide the generation process (Vechtomova et al. 2021). This means the model can produce sentences that are not only original but also relevant to a particular context or artistic intent.

6.2 Large Language Models (LLMs)

Large Language Models (LLMs) are currently among the most powerful models for generating texts. These models are trained on vast amounts of text data and are designed to predict the next word in a sequence, given the preceding context.

The core component of LLMs is the Transformer architecture, which uses the mechanism of self-attention to process all words in a sentence at once. This allows the model to understand relationships between words regardless of their position in the sentence, making it highly effective at capturing long-range dependencies in the text.

The learning objective of LLMs is typically framed as a language modeling task, where the model learns to predict the probability of the next word in a sequence. For example, given the partial sentence *The bird sat on the*, the model would assign probabilities to different possible next words (e.g. *branch*, *perch*) based on patterns it learned during training.

Once trained, LLMs can generate coherent and contextually appropriate text by sampling words one by one, using the previous words as context. Unlike the LSTM-VAE, the Transformer-based LLMs can generate longer texts.

In linguistic terms, LLMs excel at capturing surface-level regularities and common syntactic constructions but can be less effective at generating linguistically novel or unexpected expressions. This predictability can limit their utility in stimulating linguistic creativity, as the generated text may lack the unpredictability to serve as effective creative stimuli.

6.3 AI-generated lines as linguistic stimuli during the Seed Phase

I contend that LSTM-VAE models are more suitable for generating linguistic stimuli that balance randomness and predictability, making them effective tools during the Seed phase of the creative process. The key reasons are summarized below:

Controlled randomness. LSTM-VAE models introduce randomness through latent space sampling, leading to the generation of linguistically novel sentences that can disrupt habitual language patterns and stimulate creativity. LLMs do have an element of randomness in the sampling of the next token, which is controlled by a parameter that determines how likely the model is to choose less-probable words instead of always selecting the most predictable option. However, it is very limited as the model only samples from among the small number of tokens that have the highest probability of occurring in the next position.

Syntactic diversity and defamiliarization. The latent representations in LSTM-VAE models capture a wide range of syntactic variations, allowing for the generation of syntactically diverse and syntactically defamiliarized lines. LSTM-VAE samples the latent vector representing the entire sentence from the latent space, which opens up the possibility of generating dramatically different sentences syntactically and stylistically.

Limited memorization. Due to their smaller size and training objectives, LSTM-VAE models are less prone to memorizing training data, thus reducing the likelihood of generating clichéd or overly predictable language. Moreover, the training dataset can be curated by, for example, only allowing sources that conform to the desired artistic style or genre. This kind of careful curation is not feasible with LLMs, as they require training data that is orders of magnitude larger to effectively train their billions of parameters.

Transformer-based LLMs prioritize coherence and adherence to learned linguistic patterns, which may result in a text that is less effective in disrupting linguistic expectations and stimulating creativity.

7. Comparative analysis of AI-generated lines as sources of artistic inspiration

In this section, I analyze the characteristics of lines generated by two different AI systems – LSTM-VAE and ChatGPT – focusing on their potential to serve as seeds for artistic inspiration. The analysis compares two poems (Figure 1). The first poem (Vechtomova 2018) (Figure 1a) emerged through the author's creative engagement with a pool of 300-400 lines generated by an LSTM-VAE model (Vechtomova et al. 2018). Rather than following a predetermined structure or narrative, the creative process involved first selecting lines that resonated emotionally, then discovering connections and meaning through their arrangement. In contrast, the second poem (Figure 1b) was generated autonomously by ChatGPT-4o in response to a prompt requesting original lyrics.

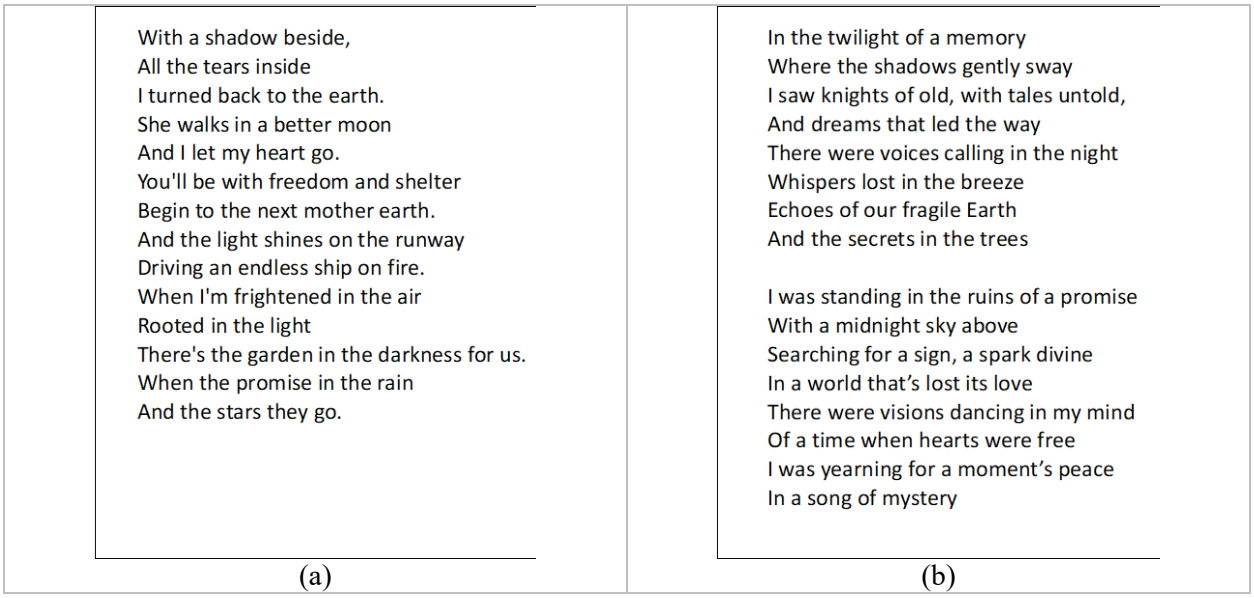

With a shadow beside,
All the tears inside
I turned back to the earth.
She walks in a better moon
And I let my heart go.
You'll be with freedom and shelter
Begin to the next mother earth.
And the light shines on the runway
Driving an endless ship on fire.
When I'm frightened in the air
Rooted in the light
There's the garden in the darkness for us.
When the promise in the rain
And the stars they go.
(a)
In the twilight of a memory
Where the shadows gently sway
I saw knights of old, with tales untold,
And dreams that led the way
There were voices calling in the night
Whispers lost in the breeze
Echoes of our fragile Earth
And the secrets in the trees
I was standing in the ruins of a promise
With a midnight sky above
Searching for a sign, a spark divine
In a world that's lost its love
There were visions dancing in my mind
Of a time when hearts were free
I was yearning for a moment's peace
In a song of mystery
(b)

Figure 1: Poems: (a) Poem #1 composed by the author from lines generated by LSTM-VAE; (b) Poem #2 generated autonomously by LLM (ChatGPT-4o)

7.1 LSTM-VAE-generated lines

The LSTM-VAE system produces lines with several qualities that make them particularly effective as creative catalysts:

Semantic openness. The lines often possess an intentional incompleteness or ambiguity that invites completion by the artist's imagination: "with a shadow beside," "all the tears inside," "rooted in the light." Such semantic openness works not through randomness but through carefully balanced indeterminacy. Each line provides enough concrete imagery or emotional resonance to anchor meaning, while remaining open enough to invite multiple interpretations.

Emotional resonance without explicit statement. Rather than directly stating emotions, the lines create emotional spaces through suggestive imagery: "when I'm frightened in the air," "and I let my heart go," "she walks in a better moon." These lines evoke emotional states while leaving their specific context undefined, allowing artists to project their own experiences and interpretations.

Unconventional juxtapositions. The system creates unexpected combinations that can spark new creative directions: "driving an endless ship on fire," "rooted in the light." These combinations challenge conventional associations while remaining emotionally evocative. The first line creates an unusual and dramatic imagery, which defies immediate resolution as it requires the reader to reconcile paradoxical combinations, such as "endless ship." The second line "rooted in the light" is evocative because it combines two concepts with different associations: "rooted" is associated with stability and permanence, while "light" is associated with something fleeting and ephemeral.

Syntactic estrangement and incompleteness. The lines exhibit both deliberate syntactic incompleteness and semantic-syntactic defamiliarization. Many lines have an open-ended quality that invites completion or continuation: "when the promise in the rain," and "and the stars they go." Such syntactic fragments do more than invite completion, they exemplify how linguistic estrangement can generate new poetic possibilities. When the text presents constructions like "walks in a better moon" or "begin to the next mother earth," it isn't simply varying its structural

patterns. Rather, these disruptions of expected syntactic-semantic relationships create what we might call *productive ungrammaticality.* The seeming *errors* (like the doubled subject in "the stars they go") function as deliberate poetic devices that slow down perception and force fresh engagement with language. These moments of syntactic estrangement transform conventional linguistic patterns into spaces of semantic possibility, where meaning emerges not from standard grammatical relationships but from more fluid, associative connections. In this way, the text's syntactic disruptions serve the same purpose as Shklovsky's *ostranenie* – making the familiar strange to heighten perception and unlock new modes of understanding.

7.2 ChatGPT-generated lines

ChatGPT produces lines that demonstrate different characteristics:

Syntactic and semantic completion. Each line demonstrates technical sophistication through complete, but conventional, syntactic-semantic relationships: "where the shadows gently sway" (adverbial clause with standard modifier-verb structure), "I saw knights of old, with tales untold" (transitive construction with parallel prepositional phrases).

While technically accomplished in their varied structures, these lines operate within expected linguistic patterns. Their adherence to conventional syntax, while creating polished poetry, actually constrains their potential for defamiliarization. Unlike texts that employ productive syntactic disruption, these complete syntactic units leave less room for the kind of creative reinterpretation that Shklovsky identified as essential to artistic perception.

Conventional imagery and tropes. The lines rely heavily on established poetic imagery: "whispers lost in the breeze," "with a midnight sky above," and "and dreams that led the way." While evocative, these images draw from familiar poetic vocabulary rather than creating new conceptual spaces.

Narrative coherence. The lines are constructed to fit into a clear narrative framework: "I was standing in the ruins of a promise," and "searching for a sign, a spark divine." This narrative clarity, while aesthetically pleasing, may limit their ability to spark unexpected creative directions.

Traditional poetic devices. The lines consistently employ conventional poetic techniques, such as hyperbaton ("spark divine," "tales untold"), metaphor ("twilight of a memory") and regular meter and rhyme. While skillfully crafted, this adherence to traditional poetic forms may constrain rather than stimulate creative exploration.

7.3 From inspiration to creation

Poem #1 (Figure 1a) demonstrates how LSTM-VAE-generated lines can serve as building blocks for original creative work. Rather than selecting lines to fit a preconceived narrative, the author's emotional resonance with certain lines guided the initial selection, after which a deeply personal narrative emerged organically through the process of engaging with and arranging these fragments. What began as an intuitive response to evocative imagery and emotional undertones gradually crystallized into a story of departure, loss, and memory. The progression from vulnerability to transcendent resolution was not planned but rather discovered through the act of composition itself. While this emergent narrative holds profound personal meaning for the author, the poem's suggestive imagery and emotional depth invite readers to form their own interpretations, demonstrating how AI-generated fragments can inspire authentic artistic creation without constraining it to predetermined paths. The poem's interpretive nature is possible due to several qualities of the LSTM-VAE lines:

- Their syntactic estrangement ("when the promise in the rain") creates productive disruptions that enable novel semantic connections.
- The paradoxical imagery ("rooted in the light") creates a strong emotional effect as they are not immediately recognizable.
- The semantic-syntactic tensions in certain images ("she walks in a better moon," "driving an endless ship on fire") open spaces for metaphorical development.
- The deliberate incompleteness of some lines ("with a shadow beside") invites readers to resolve their inherent indeterminacy.

This case study suggests that the very qualities that make LSTM-VAE lines seem less *polished* – their ambiguity, incompleteness, and unconventional combinations – may make them more valuable as seeds for artistic inspiration. While ChatGPT produces more technically accomplished lines, their completeness and adherence to convention may make them less effective as creative stimuli. This suggests that in developing AI systems as tools for artistic inspiration, we should focus less on traditional metrics of poetic quality and more on qualities that create space for human creative interpretation and transformation.

8. Conclusion

This paper set out to explore the characteristics of AI-generated poetic lines as possible stimuli for artistic inspiration during the Seed phase of the creative process. For decades, artists, writers, and poets have used methods like the cut-up technique and deliberate misuse of language to break from conventional norms and habitual uses of language. While these techniques introduced elements of randomness, they often worked within aesthetic constraints. As Nees (1969) observed, randomness alone cannot create art without an aesthetic structure to guide it. This insight is supported by studies in the psychology of aesthetics, which suggest the existence of the inverted U-shaped (Wundt) effect, according to which, artifacts of intermediate complexity are the most stimulating and aesthetically pleasing.

As neural generative models are increasingly used by people for a variety of creative tasks, they have the potential to generate stimuli that might be conducive to inspiring the artistic process. I performed a comparative analysis of two different approaches to AI-generated poetry: (a) individual lines generated by an LSTM-VAE system that were later curated by a human author into a complete poem, and (b) a complete poem autonomously generated by an LLM. This analysis revealed fundamental differences in how these systems generate content that might serve as creative inspiration.

The LSTM-VAE-generated lines demonstrate syntactic defamiliarization and indeterminacy, with their very incompleteness and ambiguity serving as catalysts for creative discovery. This is evidenced in how meanings and narratives emerged organically through the process of engaging with these lines, resulting in an original poem that transcends its individual components. The resulting work demonstrates how the semantic openness and emotional resonance of the LSTM-VAE lines enabled not just creative transformation, but the forming of an emotionally authentic artistic narrative.

In contrast, while an LLM produces technically accomplished and aesthetically pleasing poetry, its output leans toward conventional imagery and complete poetic units that leave less room for creative interpretation. The system's adherence to traditional poetic devices and narrative coherence is impressive from a technical standpoint but may limit its usefulness as a source of

creative inspiration.

This distinction aligns with Pepperell's (2007) observations about indeterminacy, which causes our habitual recognition to be suspended as the mind tries to assign meaning to the indeterminate stimulus. This is consistent with Gold et al.'s (2019) explanation of the Wundt effect, who state that the mind engages more meaningfully with stimuli in the intermediate range between randomness and predictability. The LSTM-VAE-generated lines, with their combination of structural complexity and semantic openness, appear to occupy this fertile middle-ground between chaos and convention – a space where artistic meaning can emerge through engaged creative exploration.

I contend that it is not sufficient to produce technically accomplished verses or to create superficial combinations of unrelated concepts. To serve as effective catalysts for artistic inspiration, generated lines should possess qualities that create space for creative discovery such as emotional resonance without explicit statement, syntactic openness that invites completion, and unconventional juxtapositions that spark new associations. The organic emergence of narrative through engagement with LSTM-VAE-generated lines demonstrates how these characteristics can enable artistic creation.

This suggests that in developing AI systems as tools for artistic inspiration, we should focus less on traditional metrics of poetic quality and more on generating output that occupies the fertile middle-ground between randomness and predictability. The goal should be to create systems that generate truly evocative seeds – unexpected yet meaningful, unresolved yet emotionally resonant – inviting artists to discover the narratives and meanings that emerge through deep creative engagement with these fragments.

References

Althuizen, Niek. 2021. Revisiting Berlyne's inverted U-shape relationship between complexity and liking: The role of effort, arousal, and status in the appreciation of product design aesthetics. *Psychology & Marketing* 38(3). 481-503.

BBC. 1975. *Cracked Actor. Documentary film.* (excerpt showing David Bowie discussing cut-up technique) (https://youtu.be/6nlW4EbxTD8?si=4LHeAmla2DTzN8xH) (Accessed January 24, 2025)

Berlyne, Daniel E. 1975. Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation. *Journal of Aesthetics and Art Criticism* 34(1). 86-87.

Burroughs, William S. 2012. *The job: Interviews with William S. Burroughs*. London: Penguin Books Limited.

Cross, Charles R. 2001. *Heavier than heaven: A biography of Kurt Cobain*. New York: Hyperion.

Gold, Benjamin P. & Pearce, Marcus T. & Mas-Herrero, Ernest & Dagher, Alain & Zatorre, Robert J. 2019. Predictability and uncertainty in the pleasure of music: A reward for learning? *Journal of Neuroscience* 39(47). 9397-9409.

Hochreiter, Sepp & Schmidhuber, Jürgen. 1997. Long short-term memory. *Neural Computation* 9(8). 1735-1780.

Holzer, Jenny. 1978-1987. *Truisms*. Photostat. New York: The Museum of Modern Art. (https://www.moma.org/collection/works/63755) (Accessed January 23, 2025)

Jones, Dylan. 2017. *David Bowie: A life*. New York: Crown Archetype.

Kingma, Diederik P. & Welling, Max. 2014. Auto-encoding variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, Banff, Canada.

Kruger, Barbara. 1990. *I shop therefore I am*. Photolithograph on paper shopping bag. New York: The Museum of Modern Art. (https://www.moma.org/collection/works/64897) (Accessed January 23, 2025)

Lowder, Matthew W. & Choi, Wonil & Ferreira, Fernanda & Henderson, John M. 2018. Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science* 42.1166-1183.

Molnar, Vera. 2019. *Vera Molnar: Randomness.* (https://vimeo.com/372579247) (Accessed January 23, 2025)

Nees, Georg. 1969. *Generative computergraphik.* Stuttgart: Universität Stuttgart. (Doctoral Dissertation.)

Palahniuk, Chuck. 2020. *Consider this: Moments in my writing life after which everything was different*. New York: Grand Central Publishing.

Palahniuk, Chuck. 2011. Learning from clichés… then leaving them behind. *LitReactor*. (https://litreactor.com/essays/chuck-palahniuk/learning-from-clich%C3%A9s%E2%80%A6-then-leaving-them-behind) (Accessed June 20, 2025)

Palahniuk, Chuck. 2002. She breaks your heart: Chuck Palahniuk on Amy Hempel. September 20-26. (https://www.csub.edu/~mault/palahniuk.htm) (Accessed July 1, 2025)

Pepperell, Robert. 2007. Art, perception and indeterminacy. *Contemporary Aesthetics* 5. Article 8.

Perelman, Bob. 1978. *7 Works*. Berkeley, CA: The Figures.

Rubin, Rick. 2023. *The creative act: A way of being*. New York: Penguin.

Schwitters, Kurt. 2021. *Myself and my aims: Writings on art and criticism*. Luke, Megan R. (ed.). Chicago: University of Chicago Press.

Shklovsky, Victor. 1965. Art as technique. In Lemon, Lee T. & Reis, Marion J. (eds.), *Russian formalist criticism*, 3-24. Lincoln: University of Nebraska Press.

Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan N. & Kaiser, Łukasz & Polosukhin, Illia. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* 30. 5998-6008.

Vechtomova, Olga. 2018. *Echoes I-VII*. AI Art Gallery, NeurIPS Workshop for Creativity in Art and Design. (https://www.aiartonline.com/community/olga-vechtomova/)

Vechtomova, Olga & Bahuleyan, Hareesh & Ghabussi, Amirpasha & John, Vineet. 2018. *Generating lyrics with variational autoencoder and multi-modal artist embeddings*. arXiv preprint arXiv:1812.08318.

Vechtomova, Olga & Sahu, Gaurav & Kumar, Dhruv. 2021. LyricJam: A system for generating lyrics for live instrumental music. In Gómez de Silva Garza, Andrés & Veale, Tony & Aguilar, Wendy & Pérez y Pérez, Rafael (eds.), *Proceedings of the 12th Conference on Computational Creativity*, 122-130.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:2.13-23 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

The rhetoric of the *Ukrainization of Poland* – mechanisms and effects

Grażyna Drzazga[1] & Magda Stroińska[2]
[1]*Ohio State University,* [2]*McMaster University*

Abstract: This study examines social media posts and comments reflecting changing Polish attitudes toward Ukrainians. Using thematic analysis, we identified several patterns and areas of concern. We further identified referring expressions used to describe Ukrainians in general and Ukrainian war refugees in Poland in particular. These expressions often condense opinions about individuals while revealing the speaker's attitudes. Based on previous studies on hate speech in general and in Poland, we also look at the mechanisms of expressing prejudice towards this minority group, trying to pinpoint not only main themes, but also grammatical structures used to denigrate the refugees. The unprecedented situation created by the massive wave of Ukrainian refugees seeking a safe haven in Poland and the outpouring of empathy give hope that the forced proximity of the two ethnic groups, that have shared both a good and bad political past, may help to heal the old wounds and create new bonds.

Keywords: hate speech, thematic analysis, online hate discourse, social media

Languages: English, Polish

1. Introduction

WHEN RUSSIA INVADED UKRAINE on February 24, 2022, the response of the citizens of Poland seemed unanimous: they all agreed they needed to welcome their neighbors escaping from the war zone. Photos showing the selfless involvement of Poles – who literally overnight and with minimal support from the government, built a complex system of humanitarian assistance and opened their homes to host fleeing Ukrainians – were ubiquitous in the media.

More than three years later, the situation is different. Although the majority of Polish people are still convinced their country should keep helping Ukraine and refugees, the voices of concern about the effects of the long-term conflict on the quality of life in Poland occur more and more frequently and so do the instances of hostility and hate speech towards Ukrainians. There is, for instance, a movement *Stop ukrainizacji Polski* 'Stop the Ukrainization of Poland' initiated by the far-right political party, the Confederation of the Polish Crown, whose activities seem to be gradually gaining support (Stroińska & Drzazga 2020; Drzazga et al. 2022).

In 2010, we began investigating the mechanisms of hate speech in Poland (Stroińska & Drzazga 2017, 2020; Stroińska & Cecchetto 2019; Drzazga, Stroińska & Cecchetto 2022). Fifteen years later, our data set has only grown. Thus, we shall delve into the discourse surrounding the so-called

Ukrainization of Poland, examining the mechanisms and effects of this rhetoric within a larger contexts of hate speech and its effects (Waldorn 2012).

While we are not the first to investigate hate speech directed at Ukrainians, and our analysis may not be the most extensive (Czachur et al. 2022, Liber-Kwiecińska 2023; Tomaszewska-Gałuszka 2023; Tymińska et al. 2023), it remains valuable. Grounded in our previous research, it contributes an additional perspective to the ongoing conversation. As every voice matters, we hope that our findings will resonate and be heard in the broader discussion on this critical issue.

1.1 Limitations

This study has several limitations. First, we recognize that comment sections online are moderated, which, while intended to foster a safer environment, can result in the underrepresentation of certain viewpoints. Additionally, some comments may originate from trolls or even paid bots, potentially distorting the authenticity of the discourse and skewing our findings. Lastly, by concentrating solely on publicly accessible social media, our analysis may overlook the broader context of hate speech as it occurs across various platforms. Despite these limitations, we believe our study provides valuable insights into the prevalence and nature of hate speech within this specific online space.

1.2 The roots of the term *Ukrainization*

Before we begin our analysis, let us briefly focus on the term *Ukrainization* itself. Its morphology evokes significant connotations, as it mirrors words like *Rusyfikacja* 'Russification' and *Germanizacja* 'Germanization.'

In the Polish context, these terms refer to specific historical processes from the time of the partitions of Poland (1772-1918) imposed by the neighboring countries – Prussia, Russia, and Austria – and aimed at assimilating the Polish population by imposing their own languages, cultures, and political systems. *Rusyfikacja* and *Germanizacja* were intended to erase the distinct identities of the subjected people during the time when Poland lost its independence in the 18th and 19th centuries.

Ukrainization draws on the deeply engrained negative connotations of *Rusyfikacja* and *Germanizacja* to describe the perceived cultural and demographic changes resulting from the influx of Ukrainian refugees following the Russian invasion of Ukraine. This term, its appearance and growing usage, reflects changes in attitudes and serves as a cover for much of the hostility and hateful rhetoric directed towards Ukrainians.

1.3 Polish response to the Russian invasion of Ukraine

When Russia invaded Ukraine on February 24, 2022, Poland exhibited a strong and unified response characterized by significant support for Ukrainian refugees. According to the Polish Border Guard, about 18.8 million Ukrainians, mainly women and children, crossed the Polish border between the start of the invasion and February 2024.[1]

[1] Data from: https://300gospodarka.pl/news/uchodzcy-z-ukrainy-w-polsce-liczba

Polish citizens, with minimal governmental assistance, quickly mobilized to provide aid to Ukrainian refugees. Media coverage highlighted the widespread altruistic efforts of Poles who created a comprehensive system of humanitarian support. At that time, social media were flooded with posts from influencers, celebrities, and everyday citizens sharing their personal stories of involvement. Many opened their doors to refugees, offering shelter and food, organizing fundraisers, and leading supply drives. This unprecedented wave of solidarity highlighted Poland's commitment to supporting their Ukrainian neighbors in a time of crisis.

This sentiment was widely echoed in the media with messages like *Witajcie w Polsce!* 'Welcome to Poland!,' emphasizing the warmth and hospitality extended to the refugees. The phrase *Musimy pomóc sąsiadom* 'We must help our neighbors' captured the sense of duty felt by Poles towards Ukrainians, and the slogan *Polak, Ukrainiec – dwa bratanki* 'A Pole and a Ukrainian – two brothers' highlighted the close bond between the two nations.

However, as the war continues, the attitudes of Poles are changing. First of all, Polish citizens are naturally worried about their own safety and the possible Russian response to Polish assistance to Ukraine. Furthermore, they express concerns about the alleged violence of refugees (Ukrainian and other) who live in Poland and its effect on crime statistics. Such concerns are often expressed on social media but not always supported by facts. Some Poles are also anxious about their employment opportunities, acknowledging that Ukrainian refugees are often willing to accept worse working conditions than Poles. Yet another area of concern is access to medical care; Poles are concerned that the "privileged" (as it is often referred to in these comments) status of refugees will translate into even longer waiting times to see a specialist. Prominent are also the voices expressing disappointment in the actions of the former Law and Justice government, who did not take on any responsibility for systemic assistance to refugees, leaving it to local authorities, Non-governmental Organizations (NGOs), and individual citizens.

1.4 Changes in attitudes

Public opinion surveys reveal an intriguing trend in Polish people's attitudes. While Poles' positive approach towards offering military assistance to Ukraine has remained virtually unchanged since the start of the war, a different narrative emerges regarding attitudes towards refugees.

When asked whether their views on refugees from Ukraine had changed in the past six months (since June 2022), a quarter of respondents answered affirmatively. Among those respondents, 68% reported a negative change, indicating a worsening attitude.[2]

This phenomenon has been labelled as *positive attitude dissonance,* coined to capture the contradiction in perceptions regarding the Ukrainian refugee community. On the one hand, there exists a generally positive sentiment among Poles towards Ukrainian refugees, as evidenced by various supportive indicators, including Poland's humanitarian efforts. On the other hand, the significant percentage reporting a negative shift in attitude underscores a notable discrepancy, and this change is very visible in the materials we analyzed.

[2] Data from https://wnpism.uw.edu.pl/en/further-results-of-research-on-the-social-perception-of-refugees-from-ukraine/; published in January 2023.

1.5 Stop the *Ukrainization of Poland* movement

The most prominent is the movement *Stop Ukrainizacji Polski,* 'Stop the Ukrainization of Poland,' initiated by the far-right political party, the Confederation of the Polish Crown. This group aims to counter what they perceive as an increasing influence of Ukrainian culture, values, or political interests in Poland. The movement reflects a nationalist and often xenophobic sentiment that opposes a perceived threat of cultural and political integration with Ukraine, fueled by historical tensions, geopolitical concerns, and migration issues.

The *Stop Ukrainizacji Polski* movement, which initially started as a fringe phenomenon, has been gaining momentum, particularly among nationalist groups. The movement's rhetoric frequently appeals to nationalist sentiments, leveraging fears and anxieties surrounding perceived external threats. A particularly striking example of the movement's visual rhetoric is a poster shared on Facebook, featuring the outline of Poland filled with the colors of the Ukrainian flag. In another image, also shared on Facebook, the iconic logo of Solidarność, the Polish labor union that was key in the fight against Communist rule during the 1980s, has been modified. The familiar red letters, traditionally representing the word *Solidarność* 'solidarity,' are replaced with a word that translates to *get out,* but with a much more vulgar tone.

2. Thematic analysis: Identifying main themes

In the light of these changes of attitude, we conducted our research. In our methodology, we use blogs, vlogs, videos, and users' online comments as our primary sources. Through thematic analysis, we identified and explored recurring themes, patterns, and topics across these diverse mediums, providing insights into audience interests and preferences. After the initial thematic analysis, we conducted a linguistic analysis of the data. Thematic analysis is a qualitative approach primarily used to identify, analyze, and interpret meaning patterns (themes) within a set of textual data. This method is well suited for research that aims at understanding people's views, knowledge, experiences, biases, or values based on (in our case) linguistic data (Terry & Hadfield 2021). The first theme we have identified in the collected data is **fear and concerns about safety**.

In online comments and discussions, it is not uncommon to encounter unverified stories or anecdotes depicting Ukrainians in a negative light, particularly regarding instances of violence. These stories often circulate without substantiation, fueled by allegedly personal experiences, sensationalized media reports, or preconceived biases. The examples below have been taken from the comments section on YouTube and TikTok. We have kept the original spelling and punctuation.

(1) *miałam podobną sytuację w tramwaju, tylko ,że mi ukrainiec rozwalił okulary na głowie jak poprosiłam żeby ściszył komórkę lub założył słuchawki, jak dla mnie to chołota i tyle w temacie*
'I had a similar situation on the tram, except that a Ukrainian broke my glasses on my head when I asked him to lower the volume on his phone or put on headphones. For me, it's just rude people, and that's all there is to it.'

(2) *Ukrainiec z granatem wyszedł i polaczki pouciekali hahahaha i git Ukrainiec jest w właśćicielem sklepu to nikt pod jego sklepem ryja darł nie będzie.*
'The Ukrainian came out with a grenade and the *Polaczki* ran away hahahaha and it is all good, the Ukrainian is the owner of the store, so no one will scream in front of his store.'

(3) *a jak Ukraińcy jeżdżą pijani zabijają samochodami młode dziewczyny to nikt tego jakoś nie nagłaśnia.*
'And when Ukrainians drive drunk and kill young girls with their cars, no one talks about it.'

This spread of selective negative but often unsubstantiated information contributes to a heightened sense of fear and anxiety among the Polish public, amplifying concerns about the presence of Ukrainian refugees and their impact on local communities. This fear-based narrative significantly influences public perception and fosters a divisive atmosphere.

Poles may also harbor concerns about Ukrainian **access to employment opportunities** due to several perceived factors. Firstly, there is a fear of economic competition among certain segments of the Polish population. Many Poles worry that Ukrainians, who are often willing to work for lower wages due to economic disparities between the two countries, could potentially displace them on the job market. This concern is particularly acute in sectors where wage levels are crucial for job security. Secondly, there are concerns about the impact of subsidies that employers receive. These subsidies, intended to incentivize hiring Ukrainians to address labor shortages in specific industries, may inadvertently make Ukrainian workers more attractive to employers than Polish workers, whose employment does not bring such financial incentives. This perceived advantage could lead to preferences for Ukrainian workers over their Polish counterparts, raising concerns about fairness and equal access to employment opportunities among the Polish workforce. These concerns too are often voiced through unverified anecdotes or clickbait articles. All the data comes from YouTube comments sections.

(4) *Powody zatrudniania ukraincow: – mniejsze stawki –dłuższy okres pracy bez zjadu na bazę (nawet kilka miesięcy) Dziękuję.*
'Reasons for employing Ukrainians: – lower wages – longer period of work without going back to the base (even for several months) Thank you.'

(5) *Pracowałem w firmie gdzie 80 % pracowników niewykwalifikowanych było z Ukrainy Wieczny problem Bo Polak na tym samym stanowisku zarabia więcej Bo im się należy Bo oni są większością. Ale nie zauważyli że Polacy brali nadgodziny żeby zarobić więcej.*
'I worked in a company where 80% of unskilled employees were from Ukraine. A perpetual problem. Because a Pole earns more in the same position. Because they deserve it. Because they are the majority. But they did not notice that Poles took overtime to earn more.'

Poles also express concerns about Ukrainians' **access to medical care**, driven by several perceived factors. Firstly, there is apprehension that the heightened demand from Ukrainian refugees could strain national healthcare resources, potentially resulting in even longer waiting times and reduced access for Polish citizens. Additionally, there is concern that government policies designed to support or accommodate Ukrainian refugees in the system of national (free) healthcare services could prioritize non-citizens over Polish citizens, prompting questions about equity and fairness in healthcare allocation. These concerns underscore broader debates about resource distribution and healthcare access in the context of migration influxes. This particular phenomenon is not much different than the anti-immigrant rhetoric in Canada or the US. Examples 6-8 illustrate this theme.

(6) *Wiadomo zawsze pani będzie za swoim krajem. Ja byłam świadkiem, że parka staruszków z Ukrainy wlazła do lekarza bez kolejki bo oni z Ukrainy. Dobrze że zrobiłam awanturę ale oczywiście oni z Ukrainy to mogą. Proszę być obiektywnym.*
'It's clear you'll always support your country. I witnessed an elderly couple from Ukraine skip the queue to see the doctor because they're from Ukraine. I'm glad I raised a fuss about it, but of course, they think they can do that just because they're from Ukraine. Please be objective.'

(7) *Pomagajmy, ale naszym chorym dzieciom i emerytom których nie stać na wykup leków, a później temu nieprzyjaznemu narodowi.*
'Let's help our sick children and pensioners who can't afford to buy their medications first, and only then that unfriendly nation.'

(8) *BREDNIE! POLACY mają zdychać bo się nie doczekają należnej pomocy! Wszystko darmo, natychmiast dla bande/rowców z upa/dliny!*
'NONSENSE! POLES are supposed to die like animals because they won't get the help they deserve! Everything is free and prompt for bande/rowcy from upa/dlina!'

The last example uses the term *banderowcy* 'soldiers or supporters of Bandera' – the Ukrainian leader from the time of WWII accused of collaboration with the Nazis and of atrocities against Poles, but considered a hero in Ukraine – to refer to all Ukrainians and the term *upadlina,* which is a compound consisting of *upa* (a reference to the Ukrainian Insurgent Army established in 1942 to fight against the Nazis, the Soviets, and Poles) and the ending *-dlina* from *padlina* 'carrion.' Both are referring to expressions that are evidence of hostility of the speaker.

3. Linguistic strategies in hate speech

Let us now transition from identifying the main themes to analyzing the linguistic phenomena observed in negative comments about Ukrainians. This category encompasses a range of elements, from single-word expressions – like ethnonyms – to minimizing the impacts of the war, as well as references to the past, and the use of presuppositions and generalizations.

3.1 Ethnonyms and ethnic slurs

We have noted the frequent use of the term *przesiedleńcy* 'displaced/relocated persons' instead of *uchodźcy* 'refugees.' This linguistic choice perpetuates false narratives suggesting that Ukrainians have been intentionally "relocated" to Poland as part of a deliberate strategy, rather than seeking refuge from the conflict in their homeland. We list below and analyze other ethnonyms in popular use.

Banderowcy **n** 'Followers of Stepan Bandera,' one of the leaders of the Ukrainian nationalist movement: Historically, this term refers to Ukrainian nationalists, particularly those associated with the Organization of Ukrainian Nationalists (OUN) and the Ukrainian Insurgent Army (UPA). Currently, this expression is often used in Polish discourse to criticize or demonize Ukrainian nationalism due to its role in the deaths of Polish nationals in the Polish-Ukrainian border areas during and after WWII.

Naziole 'Nazis': derived from the word *Nazi,* with an additional pejorative suffix *-ole*. A derogatory term used to dehumanize and vilify Ukrainians. This term connects to Russian propaganda that portrayed Ukrainians as Nazis, possibly due to some Ukrainian nationalists aligning themselves with Hitler against the Soviet domination during WWII.

Ukry 'Ukes': the clipping form of *Ukrainians*. It carries negative connotations and is often used in a disparaging manner. From the morphological point of view, *Ukry* is followed by the feminine/neuter plural agreement, unlike the word *Ukraińcy* that requires masculine-personal plural agreement. This is an instrument of depersonalization and dehumanization of the subject referent. The term is also intended to dehumanize Ukrainians by phonetically evoking Orcs from Tolkien's works. Orcs are portrayed as evil, mindless beings created to serve dark forces, which reinforces negative stereotypes and antagonizes Ukrainians. Interestingly, Ukrainians are using the same expression, *Orcs*, to refer to Russian soldiers (Stroińska et al. 2024).

Ukrowie 'Ukes': another informal variant of *Ukrainians,* with the suffix *-owie.* The semantic associations are as above; however, the morphological form requires the masculine-personal plural agreement and is therefore not dehumanizing.

UPAińcy: The term *UPAińcy* is a blend of *UPA* 'Ukrainian Insurgent Army' and *Ukra-ińcy* 'Ukrainians.' This compound word implies that all Ukrainians are supporters of the UPA.

Ukraińce 'Ukrainians': the change of the ending from *-y* to *-e* changes the grammatical gender category from masculine-personal to feminine-neuter, indicating the non-personal (not human) nature of the referent.

3.2 Neologisms derived from country names

Two neologisms derived from the names of Poland and Ukraine are frequently encountered in the data. A conspiracy theory circulating online claims that the Polish government is allegedly preparing its population for the potential "absorption" of Poland by Ukraine, leading to the emergence of a state known as *Ukropolin.* The roots of '*Polin* instead of *Poland*' conspiracy theory can be traced back to 1989, following Poland's first post-1945 (almost) democratic elections, where supporters of the Solidarność movement came to power. Conspiracy theorists of that time speculated about the creation of a supposed *Polin* state under the leadership of Germany and Israel (with *Polin* being the Yiddish term for Poland).

Regarding *Ukropolin,* propagandists aim to sow fear among Poles by suggesting a looming threat to their national sovereignty. This is achieved through the manipulation of news stories, taking information out of context, building on the old antisemitic and anti-Ukrainian sentiments or even fabricating evidence to suggest a gradual Ukrainian encroachment into Polish territory.

Another term that has emerged recently to refer to Ukraine is *Upadlina*. The noun *padlina* 'carrion' or 'roadkill' is combined with the abbreviation *UPA*. In Polish, this blend carries additional connotations due to its phonetic similarity to the verb *upadać* 'to fall.' This suggests that the state referred to as *UPAdlina* is a failed or fallen state.

3.3 Generalizing terms in hate speech

The use of generalizing words, such as *everyone, all, never,* or *always,* helps to spread misinformation, reinforce stereotypes, and promote discrimination and hatred against the targeted group. In the context of discussions about Ukraine, we often encounter sweeping statements such as "Ukrainians will always hate us," which generalize the attitudes of an entire population based on perceived grievances or historical events. The example below comes from a right-wing journalist and publicist, Rafał Ziemkiewicz:

(9) *Ukraińcy będą nas zawsze nienawidzieć i wszelka okazana dobroć zostanie odpłacona niewdzięcznością. Ukraina ma swoje interesy, my też, te interesy są w wielkim stopniu zbieżne i to, że tego właściwie nie rozgrywamy, to wina naszych polityków. Należy naprawić błędy i tyle.*
'Ukrainians will always hate us and any kindness we show will be repaid with ingratitude. Ukraine has its own interests, so do we, these interests are largely convergent and the fact that we do not play this fact properly is the fault of our politicians. Mistakes need to be fixed and that's it.'

These generalizations not only oversimplify complex relationships but also foster divisions and perpetuate stereotypes. They obscure the diversity of opinions and experiences within Ukrainian society, reducing individuals to monolithic entities defined by national identity alone.

3.4 Presuppositions

Presuppositions often serve to reinforce negative stereotypes and fuel societal divisions. The two examples below highlight how presupposed notions of economic competition and societal impact can influence public perception and exacerbate tensions.

(10) *Rosną niepokoje na rynku pracy. Polacy obawiają się wpływu Ukraińców*
'Unrest on the labor market is growing. Poles are afraid of Ukrainian influence.'

In the context of a discussion regarding Ukrainian refugees, this statement implies a causal relationship between the fear of Ukrainian influence among Poles and growing unrest in the labor market. It suggests that Ukrainians entering the labor market are viewed as competitors who may threaten job opportunities for Polish citizens. This presupposes a zero-sum game where the gain of one group (Ukrainians) is perceived as a loss for another.

(11) *Jak emeryci z Ukrainy mogą uzyskać polską minimalną emeryturę po zaledwie JEDNYM dniu pracy w Polsce (nawet jeśli pracują na umowę zlecenie)?*
'How can pensioners from Ukraine get the Polish minimum pension after just ONE day of work in Poland (even if they work on a mandate contract)?'

The last statement presupposes several underlying assumptions aimed at generating negative sentiments towards Ukrainians; it implies that Ukrainian pensioners exploit a loophole to quickly access Polish pension benefits with minimal effort, suggesting they are taking advantage of the system. By highlighting that pensioners can obtain Polish minimum pensions after just one day of

work, the statement insinuates that Ukrainians are benefiting without contributing sufficiently to the Polish economy or society. It indirectly suggests that such policies or practices place a strain on Polish resources or welfare systems, potentially invoking resentment among locals who may feel their benefits are being diminished or compromised. First and foremost, by asking the question *how* Ukrainians can obtain such benefits, the speaker presupposes that they in fact can get them, which is not the case. They have to fulfill the same requirements as Poles.

3.5 References to past conflicts

In our data, we observed various ways in which Poles use references to the past to express anti-Ukrainian attitudes and hostilities. A typical example is the slogan *Wołyń Pamiętamy* 'We Remember Volhynia.' Volhynia is a geographical and historical region in Central and Eastern Europe, located between Southeastern Poland, Southwestern Belarus, and Northwestern Ukraine. It has changed hands over the centuries, having been a part of the Polish-Lithuanian Commonwealth, the Russian Empire, Poland, the Soviet Union, and now Ukraine.

The topic of Volhynia is often cited in hate speech towards Ukrainians in Poland, stemming from the deep historical trauma associated with the Volhynia Massacre during World War II. This event saw the Ukrainian Insurgent Army (UPA) perpetrate mass killings of Polish civilians in 1943-1944, resulting in tens of thousands of deaths. One needs to remember that the countries of Central and Eastern Europe were the scene of many bloody national and ethnic confrontations over centuries, including the first half of the 20th century. Timothy Snyder aptly calls them *Bloodlands* in the title of his 2010 book. The narrative of Polish national victimhood and martyrdom holds significant sway in Polish national identity. References to Volhynia are sometimes used to evoke this historical suffering and reinforce a narrative of Polish victimhood and Ukrainian brutality. Far-right and nationalist groups in Poland exploit such historical events to fuel xenophobia and nationalist sentiments. By invoking these memories, they aim to provoke fear, distrust, and resentment towards Ukrainians.

Below, we highlight two specific examples of such references. Both are responses to the installation of a monument commemorating the Volhynia massacre in Southeastern Poland.

(12) *To jest SKANDAL i WSTYD Prezydencie Duda i Wysoki Sejmie – ten pomnik powinien już dawno stać tuż obok Grobu Nieznanego Żołnierza, lub w pobliżu ambasady Ukraińskiej!!!!! My Polacy pamiętamy!!!!*
'This is a SCANDAL and SHAME, Mr. President Duda and the Parliament – this monument should have been erected a long time ago right next to the Tomb of the Unknown Soldier or near the Ukrainian embassy!!!!!! We Poles remember!!!!'

(13) *Mój dziadek zawsze mówił, że Ukrainiec gorszy był od Niemca. Był partyzantem, siedział na Montelupich. Postawię świeczkę na cmentarzu w Monachium Solln. Jestem z województwa śląskiego, mam nadzieję że uda mi się tam przyjechać, Serdecznie dziękuję za pamięć tym wszystkim ludziom, którzy wreszcie pomik przyjęli.*
'My grandfather always said that the Ukrainian was worse than the German. He was a partisan, he was a prisoner in the Montelupich [Street Nazi prison in Kraków – authors]. I will put a candle in the Munich Solln cemetery. I am from the Silesian Voivodeship and I hope that I will be able to come there. I would like to thank all those people who finally accepted the monument for their memories.'

3. Conclusion

As we contemplate a future beyond conflict, amidst the ongoing war, it is crucial to envision a path for Poland and Ukraine to restore a normal neighborly relationship. This journey requires a process of reconciliation that acknowledges deep historical wounds and grievances, while emphasizing future cooperation and mutual respect.

Like the moral from Aesop's fable, *The man and the serpent*, wounds from past conflicts may be forgiven but not forgotten (Aesop n.d.). Therefore, efforts to address historical narratives, promote cultural understanding, and foster diplomatic dialogue become imperative. Ultimately, a steadfast commitment to dialogue, reconciliation, and shared prosperity will pave the way for Poland and Ukraine to forge a normalized and constructive neighborly relationship once the war is over.

Our analysis focused on the rhetoric of hostility and hate speech that we have watched rising in Poland as the Russian invasion of Ukraine continues. This is why the only data we have presented here are illustrations of negative attitudes. It is essential to acknowledge that some of these divisive comments are frequently amplified or sponsored by external sources, including Russian entities intent on fostering discord and perpetuating historical animosities. Despite these orchestrated efforts, genuine voices of support and expressions of empathy also resonate within Polish society. These voices underscore themes of solidarity, compassion, and a common humanity that transcends national borders.

References

Aesop. (n.d.). *The man and the serpent*. (https://etc.usf.edu/lit2go/35/aesops-fables/645/the-man-and-the-serpent/) (Accessed January 1, 2025)

Czachur, Waldemar & Opiłowski, Roman & Smykała, Marta. 2022. Multimodal practices of empathy and fear in Polish refugee discourse: Analysis of magazine covers. *GEMA Online Journal of Language Studies* 22(3). 63-85.

Drzazga, Grażyna & Stroińska, Magda & Cecchetto, Vikki. 2022. Different words, different worlds: The impact of linguistic labels on societal perception of refugees. *LACUS Forum* 46(3). 14-27.

Liber-Kwiecińska, Katarzyna. 2023. Emocje na celowniku–antyukraińskie komentarze w mediach społecznościowych. *Socjolingwistyka* 37(1). 217-238.

Snyder, Timothy. 2010. *Bloodlands: Europe between Hitler and Stalin.* New York: Basic Books.

Stroińska, Magda & Drzazga, Grażyna. 2020. Toxic language of contempt: The real purpose of online hate speech. *Warsaw East European Review* 10. 79-90.

Stroińska, Magda & Cecchetto, Vikki. 2019. Can there be a 'safe haven' for trauma survivors in this social media dominated world? *TRAMES* 23(2). 223-238.

Stroińska, Magda & Drzazga, Grażyna. 2017. Public and private hate speech in Poland. *Warsaw East European Review* 7. 119-134.

Stroińska, Magda & Salamatova, Svitlana & Salamatov, Volodymyr & Jayanti Jerath. 2024. Can they ever be neighbours again? Returning to coexistence after genocide. *Warsaw East European Review* XIII(1). 175-191.

Terry, Gareth & Hayfield, Nikki. 2021. *Essentials of thematic analysis*. American Psychological Association. https://doi.org/10.1037/0000238-000

Tomaszewska-Gałuszka, Izabela. 2023. *Jak manipuluje się językiem na temat uchodźców i wojny w Ukrainie.* (https://demagog.org.pl/analizy_i_raporty/jak-manipuluje-sie-jezykiem-na-temat-uchodzcow-i-wojny-w-ukrainie/) (Accessed October 10, 2024)

Tymińska, Ada & Kropal, Paweł & Sęk, Mihał. 2023. *Przyjdą i zabiorą: antyukraińska mowa nienawiści na polskim Twitterze*. *Helsińska Fundacja Praw Człowieka.* (https://hfhr.pl/upload/2023/04/przyjda-i-zabiora-raport-o-antyukrainskiej-mowie-nienawisci.pdf) (Accessed October 10, 2024)

Waldorn, Jeremy. 2012. *The harm in hate speech*. Cambridge, Mass.: Harvard University Press.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:2.24-34 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

On the success and limitations of a quantitative study of geographic and dialect distance

Sheila Embleton & Eric S. Wheeler
York University

Abstract: In previous studies, we have used digital data sets and Multidimensional Scaling (MDS) to show a strong relationship between geographical distance and dialect variation. With our English and Finnish data sets, the relationship was visual (comparing MDS plots and actual geography) and with an extensive field study of Romanian dialects, the correlation also had formal quantified measures. With correlation values about 80%, geography was clearly significant, even if there were also other factors.

The natural next step was to apply the approach to more data, particularly looking for larger and smaller geographies. We analyzed some Chinese data (extensive geography and both language and dialect variation) and some data from a small African region (Mambila, a Bantoid language group in the border area of Nigeria and Cameroon) where the language situation was varied, but the geography very "tight." We expected much weaker correlations; we found almost no correlation. The approach gave us no insight at all. There are plausible reasons for this. It is important to note that this technique, proven so fruitful in other instances, nonetheless had its limitations.

Keywords: dialectology, geographic distance, comparison method, negative results, limits to methods

Languages: Chinese, English, Finnish, Mambila, Romanian

1. The study

IT IS WELL KNOWN that dialect variation correlates with many factors, including the sociolinguistic ones of age, gender, race/ethnicity, and social class (Labov 2001; Chambers 2009). But the obvious (and one might say, traditional) explanation for dialect variation is geography: the further apart people are, the more their languages can vary (for example, Séguy 1971 is an early statement; Nerbonne et al. 2007; Szmrecsanyi 2013). How true is this? Theory-driven thinking would suggest that the more difficult it is for two groups to communicate, the more possibility there is for variation between the groups (cf. Wheeler 2007). That suggests that distance between communities, both direct and as a function of travel time or travel distance, would correlate with dialect variation. A first test of this idea, using Romanian data (Embleton, Uritescu & Wheeler

2012, 2017), demonstrated a correlation with an *r* value of about 0.80;[1] there were some small improvements when travel distance or travel time was substituted for direct distance.

Clearly, in this case, geography is an important but not exclusive correlate of dialect variation. The question arises as to whether or not the scale of distance is important. The Romanian area examined is about 250 km long, whereas the distance from Beijing to Hong Kong is about 2000 km. It is to be expected that there is "dialect variation" (more properly, "language variation") between the two Chinese locations, but does it accord with the distance?

At the other extreme, we have data from the Mambila region of Nigeria and Cameroon, where the distances are well under 100 km, but perhaps the cultural and social distances are stronger, and they override the simple geographic factors.

Using consistent approaches to the measures of geography and dialect variation, we report the quantitative correlation in these three situations and discuss some of the issues that impact the answers. Yes, geography does correlate with dialect variation, but not always in the same degree, and for good reasons.[2]

1.1 Dialect variation in England, Finland, and Romania

Over a number of years, we have used digital data sets for studying dialect variation in English, Finnish, and a selected region of Northwest Romania, known as Crișana (Embleton & Wheeler 1997a, 1997b, 2000; Embleton, Uritescu & Wheeler 2004, 2006, 2007a, 2007b). Using multidimensional scaling (MDS) in each setting, we have created pictures of the dialect variation – pictures that place regions with similar dialect variation closer together than regions with less similarity. These pictures are generated solely with linguistic data, and yet the remarkable feature is that the linguistic pictures look very much like the corresponding geographic maps of the region, although they are not exactly the same.

1.1.1 English data

For example, in the case of the English data (from the Survey of English Dialects SED, data collected 1950-1961 https://en.wikipedia.org/wiki/Survey_of_English_Dialects), we found that linguistic clusters typically represented geographic counties, and the counties spread out from North to South following fairly closely the geographic distribution of the counties. Indeed, there were some differences: the far Southwest was much closer to London (in Middlesex) linguistically than geographically, and some counties such as Cambridge had a considerable range of variation within the cluster. But still, the similarity invited comparison between dialect variation and geography (Figure 1).

[1] We applied the Mantel test using the R system (R Core Team 2020). An *r* value is a measure of correlation (the Pearson product-moment correlation coefficient). A value of 1.00 between two sets indicates perfect correlation. A value of 0.0 indicates no correlation. 0.80 is a strong positive correlation.

[2] There is a large literature on the relationship between geography and dialect variation, a literature to which we have also contributed. An effective discussion of this topic would be lengthy and would take us away from the purpose of our contribution here, which is to argue for the importance of reporting negative results.

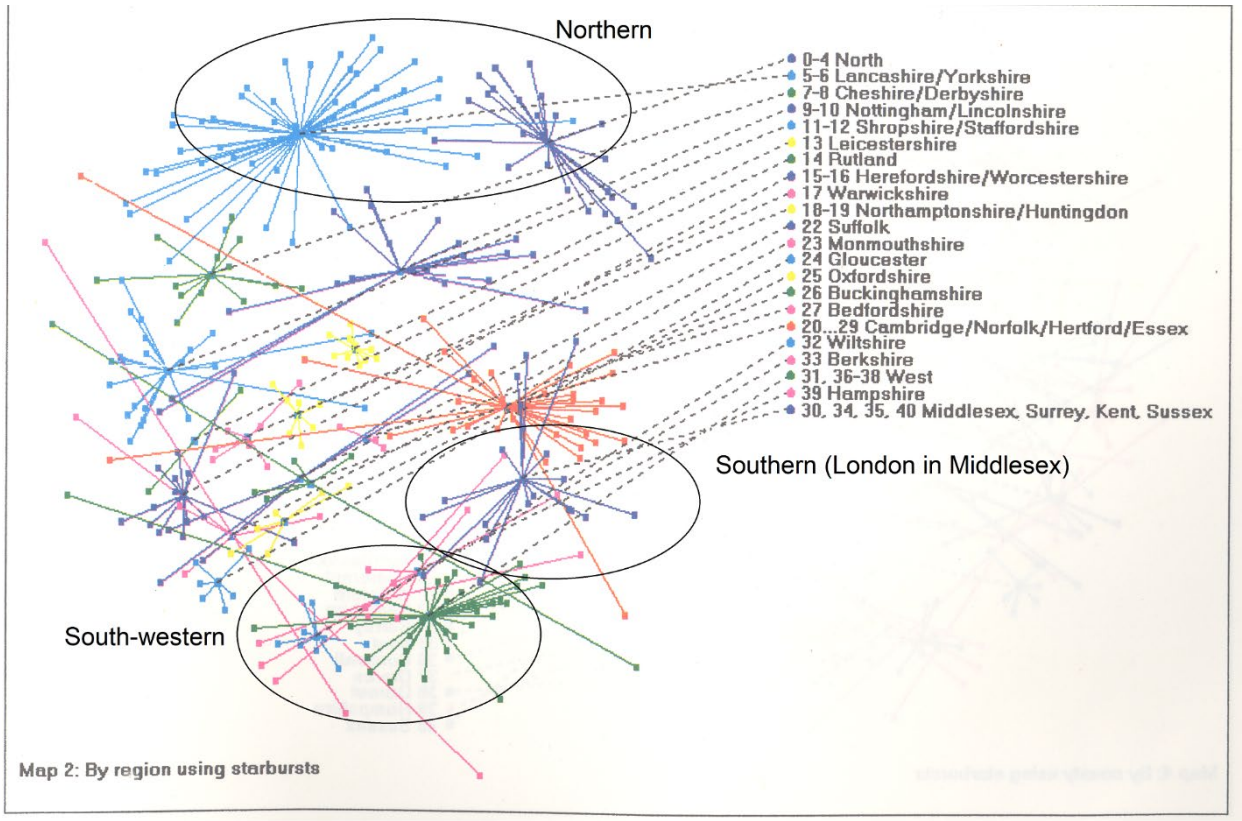


Figure 1: MDS picture of SED data showing dialect variation by groups of counties, spread from the North (at the top) to the South, including London, and Southwest (at the bottom)

1.1.2 Finnish data

For the Finnish case, we had to digitize the hard copy data from Kettunen (1940; data collected 1899-1938), and that raised a number of "editorial" issues, about which we have written earlier (Embleton & Wheeler 2000). Nonetheless, it was clear from the subsequent MDS picture that there was a "line" of dialect areas from the Southwest to the North, to the East. Again, there were exceptional situations, such as the few dialects in the West that sound like Eastern dialects, the dialect areas that were outside mainland Finland because of changing political boundaries, and the impact of Swedish on some of the dialect areas. Once more, the linguistic similarity invited an explanation based at least partially on geography (Figure 2).

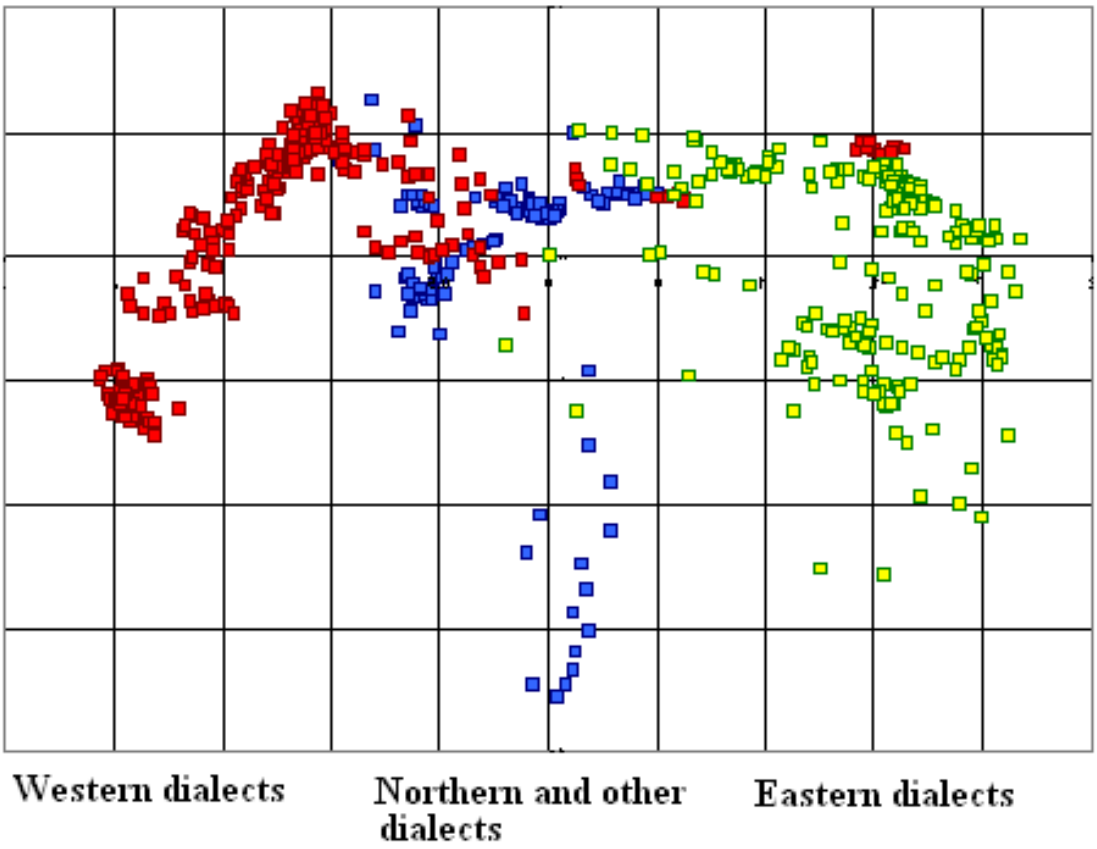


Figure 1: MDS picture of Finnish dialects showing the clear spread from West to North to East, with few exceptions

1.1.3 Romanian data

For Romanian, we were fortunate to work with Dorin Uritescu, who has created a major dialect atlas of the Crișana region of Northwest Romania (Stan & Uritescu 1996-2022).[3] The kinds of data, and the quantity of data, on dialect variation is immense. What is more, we made an online dialect atlas that gave us unprecedented access to the patterns in the data. Using built-in facilities, we could create MDS pictures based on any selection of data that we chose (Embleton, Uritescu & Wheeler 2007b).

As a result, we were able to look at dialect variation not only of *all* the data,[4] but of subsets based on phonetics, or morphology, or lexical items; the results again showed a strong similarity to geography, but interestingly, the selected subsets seemed to give a better picture of the dialect regions than using *all* the data. This raised the issue of what features should be used to define a dialect region, since the subsets did not all give the same answers. The map of phonetic variation was not the same as the map of morphological variation, and so on. Possibly merging the different subsets made the *all* picture less clear.

[3] The digitized version can be found in the York University Institutional Repository (*Romanian Online Dialect Atlas: An experimental presentation of linguistic information* (yorku.ca)). http://pi.library.yorku.ca/dspace/ (under the *dialectology* community, *RODA* collection)

[4] See Embleton, Uritescu & Wheeler (2021) for our comments on using lots of data.

However, we were able to use the distance matrices (a precursor to the MDS picture) and the R statistics package to get a quantified measure of how the dialect variation correlated to geographic distance. Although the different linguistic subsets and different definitions of geographic distance produced different correlations, they were all generally the same (ranging from 0.71 to 0.81) and generally too strong to overlook (Embleton, Uritescu & Wheeler 2012, 2017). We now felt we had "hard" quantified justification (not just an assumption) for the claim that geography was a major contributing factor to the distribution of dialect variation. It made sense: the more distant communities were from one another, the less they would influence one another. So, we set out to confirm the hypothesis in a broader range of settings.

2. Dialect variation in Mambila and Chinese

2.1 Mambila

In the border region of Nigeria and Cameroon, there are a number of communities using different (but related) language variants.[5] They have been studied by our colleague Bruce Connell (unpublished, 2000, 2001, 2006) who has provided us with a set of digitized data, from 41 locations in 17 regions, with 881 glosses based on the Swadesh list of basic vocabulary (Swadesh 1952, 1955) and other similar lists. The transcriptions are detailed, including detailed notation of tone and differences in both consonants and vowels, even for words that might be phonemically *the same*. Most items are different in some way.

In Embleton, Uritescu & Wheeler (2022), we describe how we dealt with this situation by reducing the full data strings to just consonant strings that (more or less) represent cognates. As such, we were able to get a credible multidimensional (MDS) picture of the language variation. See Figure 3.

However, when we applied the Mantel test[6] using the R software (*ncf* package (Bjornstad 2022)) for comparing the distance matrices for the geographic and linguistic distances, the correlations were very low. This is the same method used with the Romanian data. But here, the correlation coefficient was only about 0.01 indicating little or no correlation, and a p value of 0.49 indicating that even that result is not very certain. We tried various subsets of the data, and limited the data to just those records where there was some similarity between some locations. The results were no stronger. The bottom line is that in the Mambila region, one cannot predict language choices by gross geography, although it is still true that one dialect will be used in one location, another dialect in another location, and so on. Rebecca Grollemund (University of Missouri) has just started an NSF-funded project on Cameroonian languages, which will compile a large database of lexical, phonological, morphological, and syntactic data from languages of Cameroon. We have opened a conversation about possible future collaboration, so we hope to have more to report on Mambila at some point in the future.

[5] The Mambiloid languages are part of the Benue-Congo branch of the Niger-Congo language family, with about 130,000 speakers in 1993. See https://en.wikipedia.org/wiki/Mambila_language

[6] Mantel tests are correlation tests that determine the correlation between two matrices rather than between two variables.

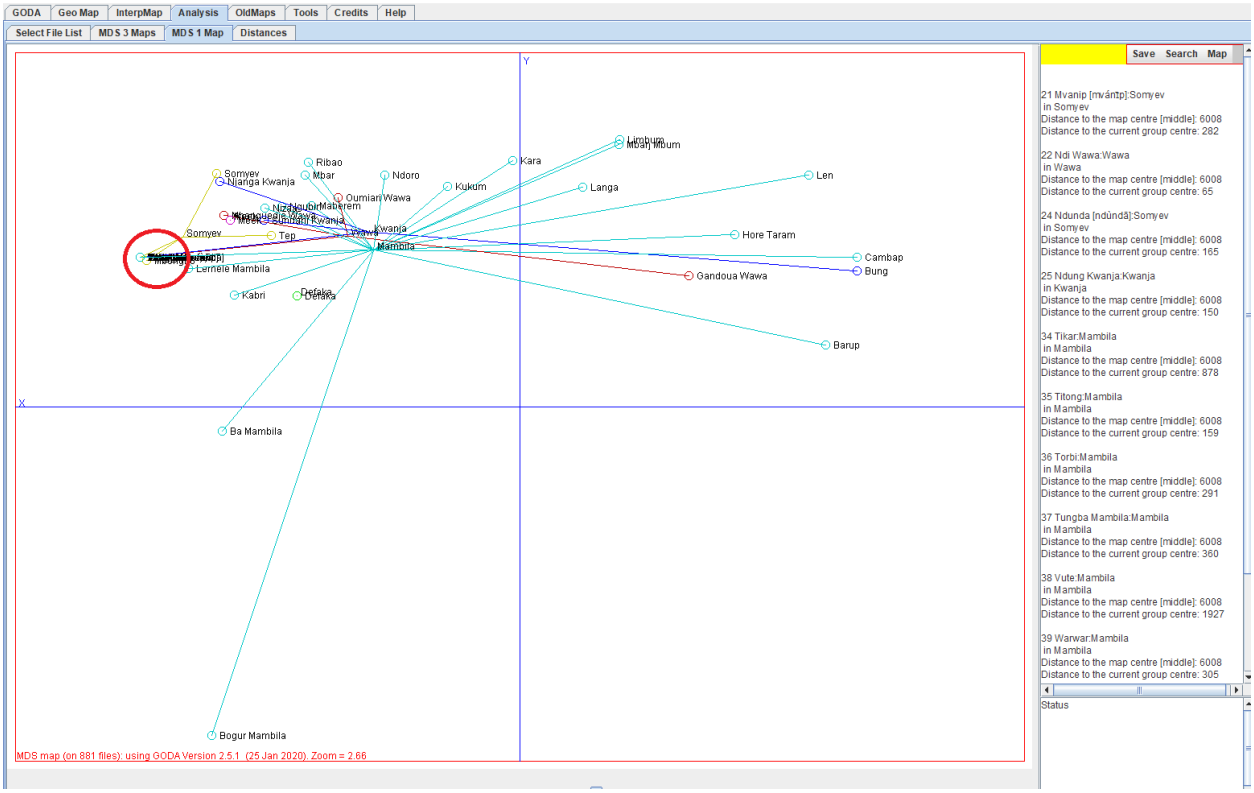


Figure 3: Mambila dialects as an MDS picture

2.2 Chinese

In collaboration with Robert Sanders (p.c.) and his various colleagues working in Beijing,[7] we have 49 sets of syntactic data for Chinese (34 regions, 930 locations, and several language varieties – some of which are orally different languages). Here we have a large geographic scale, and a reasonable expectation that there could be a geographic-linguistic correlation. However, again, the same R method indicated no strong correlation. The correlation was about 0.08, which is only marginally better than the Mambila result, and far from the 0.80 we had in Romania. Limiting the locations to one political region (e.g. Guangdong, Guangxi, Hunan, Sichuan, or Zhejiang) did not give stronger results. Although the results were surprising to us, perhaps this just shows our naïveté and lack of knowledge with respect to Chinese dialects. Tang (2018) is an overview of historical and contemporary research on the dialects of Chinese, which summarizes the field without itself contributing to the debate or applying dialectometrical methods. Ben Hamed & Wang (2006), using various quantitative methods on lexically based standardized wordlists, find that "the development of these Sinitic languages is not tree-like" (p. 29), but even network-based results "may be troublesome to interpret, especially when both the number of taxa and the number of characters increase" (p. 55). They also cite diglossia and population contact as important factors in the evolution of the Chinese dialects. List (2015), based on a dataset of lexicostatistical wordlists of length 200 from 23 Chinese dialect varieties, finds some promise in quantitative methods and network approaches as a supplement to other methods, noting the large role of contact in the

[7] Among them are researchers at the Centre for the Protection and Research of Language Resources of China, Beijing Language and Culture University, and others. A larger set of maps was not available to us because of security considerations.

development of the Chinese dialects, and the lack of scholarly agreement about details of that development. Szeto et al. (2018) also describes the complexity of the Chinese dialect situation, with internal and external contact. Even restricting their analysis to 42 varieties of just Mandarin, they can show a North-South divide, as well as much typological diversity, but also conclude that "the meticulous analysis of linguistic data, and the careful selection of language samples and features" are the "essential prerequisites for the successful application of quantitative methods" (2018: 48). Huang et al. (2024) is a comprehensive recent treatment, which reviews the history of dialect classification in the Chinese languages, and then applies dialectometrical methods to data from the *Linguistic Atlas of Chinese Dialects* (collected 2000-2008), identifying 10 major Chinese dialect regions. Again, the continuing influence of Mandarin and internal migration are found to be important factors.

3. Negative results

It is always somewhat disappointing when a promising hypothesis fails to hold in a broader context. However, it is still a valid and important result to report that a seemingly valid hypothesis has its limitations. In this case, geography does not everywhere offer a strong predictor of dialect variation. Consider that in the cases where we did see a good correlation (anecdotally in English and Finnish; quantitatively in Crișana), the geography represented a situation that could be expected to reflect communication distance:

- The communities have a common basis for variation: i.e. the English all speak English, the Finns all speak Finnish,[8] and Romanians in Crișana all use Romanian.
- The communities are physically distant but not so distant as to preclude a common communications environment (people in London do talk to people in Yorkshire, but perhaps not as often or easily as Londoners talk to one another).
- The geographic spread of dialects is long-established; one might expect that the distinctions are sustained by *tradition* in contrast, say, to differences in a rapidly evolving urban setting with high mobility and immigration where the differences might appear and disappear quickly.

If we ask, then, what is different about the Mambila setting, we see a group of communities that are all relatively close to one another, and not distributed in such a way that geography could impact the degree of communication. There is no classic gradation from one place to another as found, say, in the Romance language varieties that span the distance between northern Italy and southern France. And clearly there are strong social considerations as well. For example, Connell (unpublished) tells us that one of the varieties was used exclusively by blacksmiths. So, we might expect that geography would have no more impact on dialect here than it would in, say, New York City, where the dialect variation exists but it is not distributed by geographic distance.

On the other hand, how is the Chinese situation different? Chinese has, more or less, a common written expression, but orally it is a collection of different languages. So, to get a comparable experiment, we need to focus on, say, Mandarin. But Mandarin is, for many people, not a first language and any dialect variation may reflect the way it has been taught. Furthermore, when we apply our method to both a set of Mandarin dialects and other languages (Wu, Min), we do not get

[8] Obviously, there are also Swedish speakers in Finland, but they are a very small proportion of the population, and highly localized, including at the time of the data collection for Kettunen's atlas.

an MDS picture that shows the varieties as separate. We can claim that China does not have the right geography for our hypothesis.

However, there is another, stronger explanation to consider. In our Romanian studies, we found that different subsets of the data produced different maps of dialect variation. If we used a shibboleth based on one particular phonetic feature, we got a different distribution than if we used an alternative one (Embleton, Uritescu & Wheeler 2010, 2017). Grouping distributions according to type (e.g. phonetic vs. morphemic) gave us clear distributions, but not exactly the same distribution. We can imagine that people are aware of some of these distinctions ("Over there, they say... and we say...") but perhaps not all.[9] Still, the dialect variety is made up of many variations. Our data for both Chinese and Mambila was highly selective. In the Mambila case, there was only pronunciation data for a selected set of words. In the Chinese case, we were looking at indicators of syntactic choices, and the quantity was much smaller than in the Romanian case where we had many types of linguistic distinction, over a large data set.

Furthermore, the classic approach to the collection of dialect data (as used in the Romanian, English, and Finnish settings) focused on asking conservative, elderly speakers. In short, it represents an ideal of the way the language was spoken traditionally, possibly in a time past. In particular, the Finnish data was collected prior to 1940, and from speakers who were already old then. Perhaps this also contributed to an ideal view of dialect variation, that was not present in either the Chinese or Mambila setting.

4. Conclusion

We are left with the possibility that our negative result for the hypothesis "geography correlates with dialect variation" is the result of the need to:

- Qualify the expression "geography" to a particular range of distances and communication ease, neither too close nor too distant.
- Have a well-defined and broadly-defined definition of dialect variation.
- Have sufficient data to generate a good pattern in the language variation.

Perhaps if we defined "geography" more specifically, we might have an interesting area of study. Consider, the geography of (say) New York City, on a micro-level, defined by community, street boundaries, and so on. There are of course such studies; take as but one example Labov (1972). Then, the dialect variation from one neighbourhood to another will no doubt vary as one crosses certain boundaries... but this is really sociolinguistics, and goes well beyond the traditional notion of geography as distance.

For the geography hypothesis to make sense, there needs to be a geographic "line" or gradation along which the language can vary. It needs to be long enough to impact the ease of communication, but not so long that, say, modern technology such as telephones, internet, or air travel eliminates the effect of distance. Correspondingly, there needs to be a line or gradation of linguistic variation that can be perceived by people living along that line. The social setting must allow some communication, and there must be sufficient data (in kind and in quantity) to give identifiable dialects.

Within areas that are too large, or too small, that geographic line may not exist, or the

[9] This would take us into the realm of folk linguistics, but of course also into the literature on perceptual dialectology, e.g. the chapters in Preston (1999).

recognizable gradation of language may be missing. Both are needed, so that the one gradation can be correlated (to some degree) with the other. When these conditions exist, it seems reasonable to talk about the geographic component of language variation, but we shouldn't expect to find it in every language setting. When the conditions do exist, we have seen a high degree of correlation between geography and language variation.

5. Last comment

Quantitative linguists have mostly been supportive of our report of "negative" results, because the results were a "positive*"* limitation to the method. To some, the lack of constructive results was disappointing (essentially saying "I didn't need you to tell me there was no correlation there; I could see that already" or "Don't need a formal method to know that"), but from our point of view, this was a result or discovery about the method and it was equally important to report that negative result as the earlier results that had been positive. The method works, but only in the appropriate context of geographic and social setting, data types and quantities, and possibly other factors.

The moral of this small story then is: Listen to negative results and do not dismiss them just because there is not a positive result to point to. Every method has its successes but also its limitations, and it is important to be open about those limitations too, and not simply hide them or pretend they do not exist.

Our "shout out" to the discipline is: Report negative results.[10] Negative results are important to understanding the methods used, and thus to understanding the insight gained from our methods.

Acknowledgements

We gratefully acknowledge the ongoing support of the Social Sciences and Humanities Research Council of Canada for our research, and the contributions over many years of our late colleague and research partner, Dorin Uritescu. We are thankful to Bruce Connell for Mambila data, and Robert Sanders for Chinese data, and for their insight and comments. We also thank Ruth King for comments on an earlier version of this paper, as well as comments received from the audience at the Methods in Dialectology conference in Mainz, August 5, 2022, and at the LACUS meeting, in Waterloo, Ontario, July 2024, and from our anonymous reviewers.

References

Ben Hamed, Mahé & Wang, Feng. 2006. Stuck in the forest: Trees, networks and Chinese dialects. *Diachronica* 23(1). 29-60.

Bjornstad, Ottar N. 2022. Ncf: Spatial Covariance Functions. R package version 1.3-2. (https://github.com/objornstad/ncf)

Chambers, Jack K. 2009. *Sociolinguistic theory: Linguistic variation and its social significance*. 3rd edn. Oxford, UK & Cambridge, US: Blackwell.

Connell, Bruce. (unpublished) Data for Mambila languages. (Glendon College, York University, Toronto.)

Connell, Bruce. 2000. The integrity of Mambiloid. In Wolff, H. Ekkehard & Gensler, Orin (eds.),

[10] A Google search on "negative results" shows there are now many articles discussing this issue in science generally. One of the most accessible, laying out the issues and consequences clearly, is Kozlov (2024).

Proceedings of the 2nd World Congress of African Linguistics, Leipzig, 197-213. Cologne: Rüdiger Köppe Verlag.

Connell, Bruce. 2001. An introduction to the Mambiloid languages. In Mutaka, Ngessimo M. & Chumbow, Sammy B. (eds.), *Research mate in African linguistics. Focus on Cameroon: A fieldworker's tool for deciphering the stories Cameroonian languages have to tell; in honor of Professor Larry M. Hyman*, 79-92. Cologne: Rüdiger Köppe Verlag.

Connell, Bruce. 2006. Mambila. In Brown, Keith (ed.-in-chief), *Encyclopedia of Language & Linguistics,* 2nd ed., vol. 7, 473-475. Oxford: Elsevier.

Embleton, Sheila & Wheeler, Eric S. 1997a. Multidimensional scaling and the SED data. In Viereck, Wolfgang & Ramisch, Heinrich (eds.), *The computer developed linguistic atlas of England* 2, 5-11. Tübingen: Max Niemeyer.

Embleton, Sheila & Wheeler, Eric S. 1997b. Finnish dialect atlas for quantitative studies. *Journal of Quantitative Linguistics* 4. 99-102.

Embleton, Sheila & Wheeler, Eric S. 2000. Computerized dialect atlas of Finnish: Dealing with ambiguity. *Journal of Quantitative Linguistics* 7. 227-231.

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2004. Romanian Online Dialect Atlas. An exploration into the management of high volumes of complex knowledge in the social sciences and humanities. *Journal of Quantitative Linguistics* 11(3). 183-192.

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2006. Seeing words change using the Romanian Online Dialect Atlas. (Presentation at annual meeting of the International Linguistics Association, Toronto, April 2006.)

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2007a. Data capture and presentation in the Romanian Online Dialect Atlas. In Cichocki, Wladyslaw & Burnett, Wendy & Beaulieu, Louise (eds.), *Linguistica Atlantica* 27-28. 37-39. (*Papers from Methods XII. 12th International Conference on Methods in Dialectology.)*

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2002, 2007b. *Online Romanian Dialect Atlas*. http://pi.library.yorku.ca/dspace/ (under the *dialectology* community, *RODA* collection)

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2012. Effective comparisons of geographic and linguistic distances. In Altman, Gabriel & Grzybek, Peter & Naumann, Sven & Vulanović, Relja (eds.), *Synergetic linguistics. Text and language as dynamic systems*, 225-232. Vienna: Praesens Verlag.

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2010. Identifying dialect regions: Specific features vs. overall measures using the Romanian Online Dialect Atlas and Multidimensional Scaling. In Heselwood, Barry & Upton, Clive (eds.), *Proceedings of Methods XIII. Papers from the Thirteenth International Conference on Methods in Dialectology 2008*, 79-90. Frankfurt am Main: Peter Lang.

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2017. An expanded quantitative study of linguistic vs geographic distance using Romanian dialect data. In Wang, Lu & Köhler, Reinhard & Tuzzi, Arjuna (eds.), *Structure, function and process in texts*, 25-33. Lüdenscheid: RAM-Verlag.

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2021. The perils of big data. In Pawłowski, Adam & Embleton, Sheila & Macutek, Jan & Mikros, George (eds.), *Language and text: Data, models, information, applications*, 138-144. Amsterdam: John Benjamins.

Embleton, Sheila & Uritescu, Dorin & Wheeler, Eric S. 2022. Too much of a good thing. In Yamazaki, Makoto & Sanada, Haruko & Köhler, Reinhard & Embleton, Sheila & Vulanović, Relja & Wheeler, Eric S (eds.), *Quantitative approaches to universality and individuality in language*, 37-48. Berlin: De Gruyter.

Huang, He & Grieve, Jack & Jiao, Lei & Cai, Zhuo. 2024. Geographic structure of Chinese dialects: A computational dialectometric approach. *Linguistics* 62(4). 937-976. https://doi.org/10.1515/ling-2021-0138

Kettunen, Lauri. 1940. *Suomen murrekartasto* [The dialect atlas of Finland]. Helsinki: Suomalaisen kirjallisuuden seura.

Kozlov, Max. 2024. So you got a null result. Will anyone publish it? *Nature* 631. 728-730 https://doi.org/10.1038/d41586-024-02383-9

Labov, William. 1972. *Language in the Inner City: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2001. *Principles of linguistic change.* Vol. 2: Social factors. Oxford: Blackwell.

List, Johann-Mattis. 2015. Network perspectives on Chinese dialect history: Chances and challenges. *Bulletin of Chinese Linguistics* 8. 36-63.

Nerbonne, John & Kleiweg, Peter & Manni, Franz & Heeringa, Wilbert. 2007. Projecting dialect distances to geography: Bootstrap clustering vs. noisy clustering. (http://www.let.rug.nl/~nerbonne/papers/Nerbonne-Kleiweg-Manni.pdf)

Preston, Dennis (ed.). 1999. *Handbook of perceptual dialectology*. Amsterdam: John Benjamins.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria. (https://www.R-project.org/)

Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de linguistique romane* 139-140. 334-357.

Stan, Ionel & Uritescu, Dorin. 1996-2022. *Noul Atlas lingvistic român. Crişana.* Vol. I-V. Bucharest: Academic Press.

Szeto, Pui Yui & Ansaldo, Umberto & Matthews, Stephen. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22(2). 233-275.

Szmrecsanyi, Benedikt. 2013. *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge, UK: Cambridge University Press.

Swadesh, Morris. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *American Philosophical Society* 96. 452-463.

Swadesh, Morris. 1955. Chemakum lexicon compared with Quileute. *International Journal of American Linguistics* 21. 60-72.

Tang, Chaoju. 2018. Dialects of Chinese. In Boberg, Charles & Nerbonne, John & Watt, Dominic (eds.), *The handbook of dialectology*, 547-558. Hoboken: John Wiley.

Wheeler, Eric S. 2007. Language change in a communication network. In Grzybek, Peter & Köhler, Reinhard (eds.), *Exact methods in the study of language and text*, 689-698. Berlin & New York: Mouton de Gruyter.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:2.35-50 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Prepositions in Mixtec?
The case of gradient word classes in Tu'un Javi

DeAndré A. Espree-Conaway
University of Oregon & Bloomfield Language Institute

Abstract: This paper aims to shed light on the long-debated question of the word class of body part terms (BPTs) in Tu'un Javi – Ixpantepec Nieves Mixtec [mks], an underdescribed language spoken in Oaxaca, Mexico. Word classes in Mixtec languages are notably fuzzy, especially for spatial relator functions. I argue that BPTs contribute to the formation of prepositional constructions, exhibiting distributional behaviors distinct from their nominal origins. These BPTs and BPT phrases exhibit behavioral properties, distinct from NP possessive constructions, particularly concerning part-whole relationships. They have branched off into specific functions which distinguish them from nouns – in other words, they may have acquired syntactic behaviors, not shared with nouns. This would necessitate positing a preposition word class (or P-schema) where BPT lexemes participate in a prepositional construction – demonstrating the *leakiness* and gradience of word class formation. Tu'un Javi BPTs contribute to the body of growing evidence, supporting a Construction Grammar theoretic approach to grammatical description.

Keywords: word class, spatial grammar, fuzziness, grammaticalization, Oto-Manguean

Languages: English, Tu'un Javi (Ixpantepec Nieves Mixtec)

1. Introduction

THE NEBULOUS OR *leaky* (cf. Sapir 1921) nature of word classes is a well-known property of the world's languages, most likely reflecting the formative flexibility, necessary of an ever-changing, dynamic, complex adaptive system (cf. Schmid 2015).[1] Scholars have been aware of word class flexibility within languages across the four traditional, functionally specialized *core* word classes – that is, word classes for *referring expressions* or nouns, *predicating expressions* or verbs, and *modifying expressions* or adjectives and adverbs – for over 100 years (Rijkhoff & van Lier 2013). However, the centrality of certain word classes to human language and the extent of their gradability is still heavily debated. While many scholars are comfortable with there being

[1] Data Abbreviations: 1 = first person; 2 = second person; 3 = third person; AML = animal; CAUS = causative; COM = completive aspect; CON = continuative aspect; DEM = demonstrative; DET = definite determiner/definite specifier; FEM/F = feminine; INAN = inanimate; IND = indefinite determiner; IPFV = imperfective; HON = honorific; M = masculine; NEG = negative; PL = plural; Q = question particle; SG = singular; STAT = stative aspect.

arguably no class of adjectives in some languages and of adverbs being a *catch-all* category, many hold to the centrality of a noun–verb distinction. Despite this, other scholars have proposed that many languages do not make this distinction (namely, Iroquoian, Nootkan, Philippine, Polynesian, Salishan, and some Austroasiatic languages) (Jacobsen 1979; Croft 2001; Evans & Osada 2005). In this paper, we examine the case of adpositions.

The status of adpositions (and their many types, i.e. prepositions, postpositions, circumpositions, and ambipositions) is notoriously fuzzy, with some scholars treating them as "functional items, as lexical items, as split items and/or as semi-lexical items." (Deacon 2014) Whether adpositions constitute a word class on par with other *core* word classes, like nouns and verbs, is an ongoing debate (Chomsky 1970, 1981; Grimshaw 1991, 2005; Baker 2003; DeLancey 2005; Hagège 2010; Deacon 2014; Frajzyngier 2022). This paper aims to shed light on the question of syntactic category gradience and rigidity by examining the case of the long-debated noun/preposition (N/P) distinction in Oto-Manguean languages. The main problem centers around the fuzziness of body part term (BPT) spatial relators. For example, in (1a) and (1b) below, the anatomical BPT in a part-whole NP possessive construction has the same form as that which would receive a locative reading.

(1) a. *sàtà* *ósó*
back bear
'the back of the bear/the bear's back'

b. *sàtà* *ósó*
back bear
'behind the bear'

The similitude of these syntactic forms makes it difficult to determine whether they are nouns in both their anatomical and extended uses (i.e. relator nouns) or whether they should be considered part of the small word class of prepositions proper for their spatial and non-spatial, non-anatomical uses. The purpose of this study is to analyze the use of Tu'un Javi (Ixpantepec Nieves Mixtec [mks]) BPTs from a functional/cognitive-based, Construction Grammar theoretic perspective (Goldberg 1995, 2006; Croft 2001), leading to insights on the fuzziness of this word class.

The fundamental research question is thus: are adpositions on par with other word classes, such as one of the major word classes from which they are often (historically) derived, namely nouns? In the case of Tu'un Javi, I argue that these BPTs are implicated in the formation of prepositional constructions, based on the distributional behaviors which they do not share with their word class of origin. These BPTs and the BPT phrases in which they participate exhibit behavioral properties, not found in either basic NP possessive constructions, including those that implicate part-whole relationships. Here, we take a Construction Grammar theoretic perspective (Goldberg 1995, 2006; Croft 2001) on word class – that is, rather than ascribing certain lexemes to particular classes, various lexemes participate in word class schematic constructions (i.e. N-schemas, V-schemas, and, as I argue here, P-schemas) (Diessel 2020 and see Croft 2001; Goldberg 2006). Within this approach, as Diesel (2020) writes, "one can […] think of word classes as slots of constructional schemas."

1.1 Construction and word class

Here *construction* is defined as any form-function pairing on the syntax-lexicon continuum that meets the criteria of non-compositional grammar or high frequency. Simply put, I rely on Goldberg's (2006) definition of *construction* which implicates both novel form–function pairs with emergent functions as well as a high frequency of use, as she writes:

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (p. 5)

Due to semantic fitness and high-frequency use, BPT constructions behave like the narrow class of *true* prepositions in this language, forming functionally motivated constructions of both high-frequency and non-compositional meaning (i.e. projected space and non-spatial, relational functions, including syntactic dependencies and semantic relations). As Croft (2001: 4) writes, "constructions are the basic units of syntactic representation, and categories are derived from the construction(s) in which they appear."

This paper describes how the distributional behaviors, associated with BPTs in P-schemas, differ from their use in N-schemas, especially those that constitute NP possessive constructions (including those that express part-whole relationships). This is key to establishing that these BPT constructions are not mere relator nouns within an NP part-whole possessive construction which have extended its use while maintaining the same syntactic form. In addition to using several language-family wide diagnostics (Hollenbach 1995; Lillehaugen 2003, 2006; Munro 2006; Lillehaugen & Sonnenschein 2012), some innovations of this study are other diagnostics which I hope will facilitate cross-linguistic comparison across the Oto-Manguean language family. In addition to addressing issues in word class and syntactic theory, this paper also aims to provide more data for the underdescribed language, Tu'un Javi, spoken in Oaxaca, Mexico, and among the diasporic community around the United States.

In this paper, I begin by introducing key issues in word class typology and how the case of the N/P debate in Mixtec might shed new light on these issues (Section 1). Section 2 briefly introduces basic Tu'un Javi grammar, focusing on aspects the reader needs to know to understand the rest of the paper. In Section 3, I briefly review the word class typology of prepositions and relator nouns while laying out the arguments in Oto-Manguean studies, concerning the status of these BPTs with spatial and other relational functions. Section 4 analyzes the behavior of BPT constructs, revealing the constructions in which they participate and, therefore, emergently form. Finally, Section 5 concludes the paper, summarizing the arguments put forth along with describing the greater implications of the study.

2. Tu'un Javi: Basic typological profile

2.1 Basic grammar

Like all Mixtec languages, Tu'un Javi is a tonal language with a basic word order that we can consider predicate-initial (VSO). Tu'un Javi is a language variety of the Southern Mixteca Baja dialect cluster (Oto-Manguean family) spoken in Oaxaca, Mexico.

As for the word order, many other permutations are possible for the sake of changes in pragmatic focus. This includes *wh-* questions, replies to *wh-* questions (i.e. such as the basic locative construction), and statements of emphasis. Both S and O may be fronted before the V either individually or together. Both BPT phrases and PPs have a tendency to be fronted compared to other phrase types, however, as we shall see, NPs – including NP possessives – may also be fronted. Here, in Figure 1, I provide a list of Tu'un Javi BPTs and prepositions.

BPTs	Translation	Extended meanings	# in Bowped	# total in corpus
nùùn	'face'	'on (x),' 'at (x),' 'to (x),' 'through (x),' 'from (x),' 'in front of (x),' 'the front part of,' 'than (x),' 'where [relativizer] (x).'	25 (35%)	361
tyìsi	'belly'	'under (x),' 'bottom of (x),' 'bottom part of (x).'	4 (5.6%)	20
ini/iñi	'insides'	'in (x),' 'inside of (x).'	12 (16.8%)	196
sàtà	'back'	'behind (x),' 'on (x).'	8 (11.2%)	63
sà'à	'foot'	'under (x),' 'at the foot of (x),' 'on behalf of (x),' 'about (x),' 'for (x),' 'for the sake of (x).'	2 (2.8%)	98
kà'à	'bottom/buttocks (polite)'	'at the bottom of (x).'	0 (0%)	15
nda'a	'hand/arm'	'on (x).'	2 (2.8%)	47
yu'ù	'mouth'	'edge of (x),' …+ *nùùn* 'instead of (x).'	3 (4.2%)	110
xinì	'head'	'top of (x).'	7 (9.8%)	54
ndokó	'waist'	'around (x).'	3 (4.2%)	11

Prepositions	Translation	Uses	# in Bowped	# total in corpus
mà'ni/mà'ñi	'center/between'	'between (x),' 'at the center of (x).'	0 (0%)	3
tyání	'between'	'between (x).'	0 (0%)	1
xí'ín/xíín	'next to'	'next to (x),' 'beside (x)	1 (1.4%)	17
si'in	'with'	'with (x),' 'to [as in 'speak to'] (x)'	0 (0%)	97
vátyí	'close to'	'close to (x).'	0 (0%)	3
tátú'ùn	'like'	'like (x),' 'as (x).'	0 (0%)	79
púntá	'point'	'at the peak of (x),' 'at the head of (x).'	0 (0%)	2
ndyì	'until'	'until (x),' 'to/toward (x),' 'from (x).'	0 (0%)	41
no spatial relator	'no spatial relator'	'at (x),' 'in (x).'	8 (11.2%)	8

Figure 1: BPTs & prepositions: Forms, functions, and frequencies

3. Word class fuzziness and BPT syntax in Mixtec

From a Construction Grammar theoretic perspective, a word class is an emergent category derived from the constructs that participate in it. Several scholars have attempted to provide an overarching, typological description of adpositions (Vandeloise 1991; DeLancey 2005; Dryer 2007; Hagège 2010; Givón 2021; Frajzyngier 2022). Prepositions are defined here as P-schema or constructions in which such expressions mediate functional relationships between their head lexeme's complement and some other linguistic element in the clause. Adpositions often take their origins in relator nouns and serial verbs as shown in Figure 2.

Nouns Verbs	| Relator nouns | Serial verbs	>	adpositions	>	clitics	> cases

Figure 2: Grammaticalization pathways of spatial relators

Scholars in Oto-Manguean languages have argued for various degrees of grammaticalization of BPTs. Many have argued that BPTs are (relator) nouns. A few scholars have suggested that BPTs are "emergent prepositions." Hollenbach (1995) and Lillehaugen (2003, 2006), among others, have argued that BPTs are prepositions proper.

4. NP vs. PP: The word class status of body part terms

As mentioned previously, word classes in Mixtec languages are notoriously fluid, with words that serve spatial relator functions exhibiting some of the most ambiguous properties. However, there are several behaviors that markedly distinguish BPTs from the prototypical Tu'un Javi possessive construction, including possessive constructions that reflect part-whole relations as they function within prepositional constructions. While some nominal properties have been retained, several key

diagnostics reveal certain distinctive behaviors.

To restate the problem, because of flexible word class distinctions, flexible word order, and a lack of robust nominal morphology, morphosyntax is often unhelpful for distinguishing between the construction of the preposition and complement [P + NP] and the head noun or possessum (the *part* in part-whole structures) and the dependent noun complement or the possessor (the whole) [$NP_{POSSESSUM}$ + $NP_{POSSESSOR}$/$NP_{(PART)}$ + $NP_{(WHOLE)}$]. Example (2) provides these structures in parallel.

(2) a. **P** + **NP**
si'in *Sárá*
with Sara
'with Sara'

b. **$NP_{POSSESSUM}$** + **$NP_{POSSESSOR}$**
tòyiká *Sárá*
basket Sara
'Sara's basket'

c. **$NP_{(PART)}$** + **$NP_{(WHOLE)}$**
íxì *Sárá*
hair Sara
'Sara's hair'

Predictably, these structures with BPTs (i.e. BPT + NP) can lead to the kinds of ambiguities that we observed in example (1) above and again in example (3) below.

(3) a. **BPT** + **NP**
$NP_{POSSESSUM}$ + **$NP_{POSSESSOR}$**
$NP_{(PART)}$ + **$NP_{(WHOLE)}$**
nùùn *Yesénià*
face Yesenia
'Yesenia's face'

b. **BPT** + **NP**
P + **NP**
nùùn *Yesénià*
face Yesenia
'in front of Yesenia'

c. **BPT** + **NP**
P + **NP**
nùùn *Yesénià*
face Yesenia
'[give] to Yesenia'

The BPT + NP construct in (3a) gets a clear possessum-possessor/part-whole reading, but (3b) and (3c) get a spatial relator and dative reading respectively. While these are rather clear semantic extensions, it is not at all clear that the syntax of these structures differs. The diagnostics offered in this section – those of complement fronting or BPT strandedness, adjunction, constituency grouping, coordination, possession, quantifier modification, attributive modification, and determiner use – provide evidence of BPTs, participating in prepositional constructions.

4.1 Strandedness

Here we have the possibility of complement fronting or a *strandedness test*, a movement diagnostic described in Hollenbach (1995). While it is normal in Mixtecan languages for the whole constituent to be fronted, placing that constituent in pragmatic focus – whether it is possible to place only the complement of the BPT in this focus position and allow the spatial relator itself to

remain *stranded* – has yet to be thoroughly tested. This would not be possible of a possessor-possessee relationship and therefore would provide evidence of distinctive syntactic behaviors. As previously mentioned, fronting the entire PP phrase is possible, as demonstrated in examples (4a) and (4b) below. Example (4c) demonstrates that a *true* P can be stranded with the complement fronted to the focus position.

(4) a. *kà'àn=in* ***si'in*** ***ñá'à*** ***klárá***.
CON.talk=1SG with FEM Clara
'I was talking with Clara.'

b. ***si'in*** ***ñá'à*** ***klárá*** *kà'àn=in.*
with FEM Clara CON.talk=1SG
'I was talking with Clara.'

c. ***ñá'à*** ***klárá*** *kà'àn=in* ***si'in***.
FEM Clara CON.talk=1SG with
'I was talking with Clara.'

Now, let us take a look at BPTs in a similar configuration. Normally, a BPT phrase occurs after the VS elements of the clause as in (5a), or the entirety of the BPT phrase is fronted to the focus position preceding the VS elements. However, the possibility of leaving the BPT *in situ* while fronting its NP complement to the left, focus position also occurs (5b).

(5) a. *nì-chi+kàà=ra* *tòyiká pérá* ***nùùn*** ***bisiklétá=ra***.
COM-touch+sit.down.SG=3SG.M basket pear face bicycle=3SG.M
'He put the basket of pears on his bike.'

b. ***bisiklétá=ra*** *nì-chi+kàà=ra* *tòyiká pérá* ***nùùn***.
bicycle=3SG.M COM-touch+sit.down.SG=3SG.M basket pear face
'He put the basket of pears on his bike.'

Using that same example clause, we can see that the direct object *tòyiká pérá* 'basket of pears' – forming a typical NP possessive construction – may be fronted in its entirety (6a-b), but it may not split the head from its dependent (6c-d).

(6) a. *nì-chi+kàà=ra* ***tòyiká pérá*** *nùùn* *bisiklétá=ra.*
COM-touch+sit.down.SG=3SG.M basket pear face bicycle=3SG.M
'He put the basket of pears on his bike.'

b. ***tòyiká pérá*** *nì-chi+kàà=ra* *nùùn* *bisiklétá=ra.*
basket pear COM-touch+sit.down.SG=3SG.M face bicycle=3SG.M
'He put the basket of pears on his bike.'

c. ****pérá*** *nì-chi+kàà=ra* ***tòyiká*** *nùùn* *bisiklétá=ra.*
pear COM-touch+sit.down.SG=3SG.M basket face bicycle=3SG.M

d. ****pérá*** *nì-chi+kàà=ra* *nùùn* *bisiklétá=ra* ***tòyiká***.
pear COM-touch+sit.down.SG=3SG.M face bicycle=3SG.M basket

This splitting limitation suggests that the relationship between the BPT and its complement is not

of the same kind as the relationship between the head and its dependent in a typical possessive construction. However, because *tòyiká pérá* 'basket of pears' would not form the same type of NP possessive construction – that is, one reflecting part-whole relations, rather than a nominal modifying one as in this case – this raises the question of the behaviors of part-whole NP possessives. Between examples (7a) and (7b), we observe that the complement of the part *jò'o* 'ear' may not be stranded while the complement (whole) *loosó* 'fox' is fronted.

(7) a. *nì-chi+kàà Benísió tójò'o livì* ***jò'o*** ***loosó***.
COM-touch+sit.down.SG Benicio earring beautiful ear fox
'Benicio put the beautiful earrings on the fox's ears.'

b. ****loosó*** *nì-chi+kàà Benísió tójò'o livì* ***jò'o***.
fox COM-touch+sit.down.SG Benicio earring beautiful ear

In situations like this, it appears that the ON is recoverable from the pragmatic context since we have the presence of a part – that is, the ON is most likely, simply indexed rather than directly encoded in the utterance. According to Frajzyngier (2022), this kind of functional non-coding is a typical behavior of adpositions. He suggests that:

> [o]nce the speaker has decided to express some function encoded in the grammatical system, the speaker scrutinizes whether the lexical items carry the meanings of the selected semantic function. If the lexical items do carry the features of the semantic function chosen, no other means must be deployed to code that function. If the lexical items do not carry the features of the semantic function chosen, some grammatical markers must be deployed. (p. 148)

Also, notice how *ear* here would be compositionally locative rather than exhibit the non-compositional locative flexibility seen in other BPTs. However, what is most important is that not all BPTs may be employed within the prepositional construction and therefore only certain entrenched (high frequency) and conventionalized (non-compositional participating) BPTs may be used, forming a closed class of prepositions, typical of prepositions typologically.

BPTs, entrenched and conventionalized, in prepositional constructions like those listed in Figure 1 above may be stranded as illustrated in example (8) below.

(8) a. ***nana=ñá*** *sáku=ñá* ***sà'à***.
mother=3SG.F CON.cry=3SG.F foot
'She is crying for her mother.'

b. ***vídrió*** ***yó'ó*** *uùn kanakàà kityo'o* ***ini***.
glass here NEG be.there 3SG.AML insides
'He (the frog) didn't stay inside the glass.'

c. ***yùtyá*** *nì-já-kàkà=in* ***ndyì***.
river COM-CAUS-walk=1SG until
'I drove up to the river.'

4.2 Adjunction

BPT phrases may be used as adjuncts in intransitive clauses, just as prepositions may. For example, with the verb *kata* 'sing,' the BPT phrase with *ini* 'insides' and *kárró* 'car' as its complement may be used as an adjunct; however, *kárró* alone, despite it arguably having a value of place, is not permitted.

(9) a. *kata* *Sárá.*
CON.sing Sara
'Sara is singing.'

b. *kata* *Sárá* *ini* *kárró.*
CON.sing Sara insides car
'Sara is singing in the car.'

c. **kata* *Sárá* *kárró.*
CON.sing Sara car
≠ 'Sara is singing in the car.'

This suggests that *ini* 'insides' is not quite of the same syntactic category as its constituent *kárró* 'car' and therefore does not form the NP+NP structure (or in this particular case, simply N+N structure) that characterizes the NP possessive construction.

4.3 Constituency group

If one can substitute the complement of the BPT (with the demonstrative *kàà* 'that'), but not the possessum of an NP, then this would be good evidence against inheritance between the NP possessive construction and the construction in which BPTs participate. Example (10) demonstrates this. Notice how the complement of the BPT can be replaced by the demonstrative, but the complement of the NP possessive cannot.

(10) a. *nùùn* ***bisiklétá=ra***
face bike=3SG.M
'on his bike'

b. *nùùn* ***kàà***
face DEM
'on that'

c. *tòyiká* ***pérá***
basket pear
'basket of pears'

d. **tòyiká* ***kàà***
basket DEM
'that basket' and not 'basket of that'

e. *jò'o* ***loosó***
ear fox
'ear of the fox' or 'fox's ear'

f. **jò'o* ***kàà***
ear DEM
'that ear' and not 'ear of that'

The examples above suggest that the relationship between the BPT and its complement does not share the same syntactic status as the typical $NP_{POSSESSUM}$+$NP_{POSSESSOR}$ construct or the $NP_{(PART)}$+$NP_{(WHOLE)}$ construct. It does, however, suggest similar relations between both kinds of possessives which we might expect. This further suggests a deeper distinction: a common compositionality for both typical NP possessives ($NP_{POSSESSUM}$+$NP_{POSSESSOR}$) and their part-whole counterparts ($NP_{(PART)}$+$NP_{(WHOLE)}$) within the greater NP possessive construction – which is non-compositionally *possessive* while only an entrenched and conventionalized set of BPTs may participate in this other construction which is non-compositionally *locative*.

Now, let us examine the BPT as the ground phrase of a clause. If the ground phrase cannot be fully replaced with the *kàà* demonstrative, this might also suggest a word class difference.

(11) a. *nì-chi+kàà=ra* ***tòyiká pérá nùùn bisiklétá=ra***
COM-touch+sit.down.SG=3SG.M basket pear face bike=3SG.M
'He put the basket of pears on (the front of) his bike.'

b. *nì-chi+kàà=ra* ***kàà nùùn kàà***
COM-touch+sit.down.SG=3SG.M DEM face DEM
'He put that on that/there.'

But not,

c. **nì-chi+kàà=ra* ***kàà kàà***
COM-touch+sit.down.SG=3SG.M DEM DEM
'He put that on that/there.'

Since the BPT is required as shown between examples (11b) and (11c), this would suggest that the status of an NP like *basket of pears* differs from the status of the phrase *on his bike* – in other words, a *face of this bike* reading would not be possible here, thus placing a common word class designation in doubt.

4.4 Coordination

The possibility of coordination may be used to verify the syntactic status of BPT phrases compared to NP possessives and PPs. The PP and the BPT phrase may be coordinated by means of conjunctive coordination, as demonstrated in examples (12a) and (12b) below.

(12) a. *nì-sino tyií tyìsi mésá ta ndyi'i xíín tyayì.*
COM-run mouse belly table and then next.to seat
'The mouse ran under the table and by the chair.'

b. *ndyáa kà'à kìji tyi jani tyání nda'a vàli.*
stick bottom pot and also between hand/arm little
'It (the mud) is at the bottom of the pot and between my fingers.'

Coordination using the *a ... a* 'either … or' correlative conjunction construction is permitted between BPT phrases and PPs.

(13) a. *ku=nee kityi sikwàá a nùùn kává a púntá yukùyì'a.*
STAT=hang animal night either face cave or top volcano
'The bat is (hanging) either in the cave or at the top of the volcano.'

The PP and the BPT phrase may be coordinated by means of disjunctive coordination, as demonstrated in the following examples (14a) and (14b).

(14) a. *a ku=núun vásó mà'ni ín mésá a yu'ù mésá?*
Q STAT=be.on cup middle.of IND table or mouth table
'Was the cup in the middle of the table or on the edge of the table?'

b. *a sita Benísió sàtà nùùn xìko=ná a mà'ni*
Q CON.sing Benicio back face sell=3PL or middle.of

nùùn ityi si'in veñu'u.[2]
face cornfield with/and church
'Is Benicio singing behind the market or between the cornfield and the church?'

This data demonstrates that BPT phrases and PPs likely share the same syntactic categorical status. More importantly, both BPT phrases and PPs share this property neither with typical NP possessives ($NP_{POSSESSUM}$+$NP_{POSSESSOR}$) nor with the part-whole constructs ($NP_{(PART)}$+$NP_{(WHOLE)}$), as demonstrated in examples (15a) and (15b) below.

(15) a. **ku=ndyityi sápò tyìsi síyá tyi sà'=in.*
STAT=stand toad belly chair and foot=1SG
≠ 'The toad is under the chair and on my foot.'

b. **na-kàà valóòn tyìsi síyá tyi kájá.*
IPFV-sit.down.SG ball belly chair and box
≠ 'The ball is under the chair and in the box.'

Notice that despite *kájá* 'box' being arguably of inherent locative value, the status of this noun compared with the BPT and its phrase differs. Even with common BPTs such as *sà'à* 'foot' which is in our list above, when used in an explicitly anatomical way – and therefore participating in an N-schema – may not be coerced into a construction type that is of equivalent status to the BPT phrase (i.e. there with *tyìsi* 'belly'). It should also be noted that this points to some degree of gradience among entrenched and conventionalized BPTs and their proclivities to participate in PP constructions. While *sà'à* 'foot' is on our list, *foot* is one of the most commonly entrenched and conventionalized BPTs throughout the Otomanguean family and throughout the Mesoamerican sprachbund. In Tu'un Javi, it carries a somewhat lower frequency compared to some other BPTs which may account, in part, for this prohibition. This suggests that gradience in frequency may yield gradience in functional reach and extensiveness.

4.5 Possession

Possession generally does not allow for us to observe the kinds of distinctions that there may be; however, one possessive construction does reveal some interesting results. In the context of honorific-related NP possessive constructions, the honorific possessive enclitic, in this case, *ná* '3rd person singular masculine honorific,' is used in conjunction with the proper noun (example 16b). This is normally prohibited with either the proper noun, as in (16a), or simply the pronominal

[2] In this example, the *nùùn* 'face' is taking on its nominalizer meaning 'the place of (x)' or 'the place where (x) [is/occurs].' In this example, *nùùn xìko* 'market' is literally 'the place of selling' and *nùùn ityi* 'the place/location of the cornfield.'

enclitic.

(16) a. **CONTEXT:** Benicio is a contemporary and does not warrant honorific use.
nana *ká'ni* *Benísió*
mother big Benicio
'Benicio's grandmother'

b. *nana* *ká'ni=ná* *Benísió*
mother big=3SG.M.HON Benicio
'[The Elder] Benicio's grandmother'

In the case of example (16b), the honorific pronominal enclitic functions as a genitive marker on the head, indexing the dependent. This is not possible with BPTs in a prepositional construction.

(17) a. **nùùn=ná* *Benísió*
face=3SG.M.HON Benicio
≠ 'on/in front of [the Elder] Benicio'
= somebody else's (i.e. an elder's) face of whom that person is related to Benicio [e.g. Benicio's grandfather's face].

This difference, as a minor instance of case marking, suggests that BPT phrases do not share the same nominal morphological behaviors as NPs or NP possessives.

4.6 Quantifiers

Quantifiers occur before their constituent nouns, as demonstrated in example (18a). However, in the same way that quantifiers cannot modify prepositions, example (18c) is infelicitous because BPTs may not be modified by quantifiers when in this construction.

(18) a. *nì-seen* *Ariáná*[3] *ùnì* *mésá.*
COM-buy Ariadna three table
'Ariadna bought three tables.'

b. *nì-chi+kojo* *Ariáná* *kò'ò* *nùùn* *ùnì* *mésá.*
COM-put+sit.down.PL Ariadna bowl face three table
'Ariadna put bowls on the three tables.'

c. **nì-chi+kojo* *Ariáná* *kò'ò* *ùnì* *nùùn* *mésá.*
COM-put+sit.down.PL Ariadna bowl three face table

Note that only the NP complement of the BPT is modifiable by a quantifier. The incongruency between examples (18a) and (18c) suggests different word class categories between BPTs and canonical referencing nouns.

[3] As an aside, this should be noted as an (occasional) case of loanword adaptation: from *Ariádná* [ʔaɬriɬadˈnaˈ] → *Ariáná* [ʔaɬriɬaˈnaˈ].

4.7 Attributive/property concept modification

Whether Mixtec languages have adjective word classes is very much so still deeply debated, and Tu'un Javi is no exception to this. However, we may still speak of lexemes with primarily an attributive function or property concept words (PCWs) (Thompson 1988; Penner 2020). Many of these PCWs are predicative and are lexified in Tu'un Javi as stative verbs. Here, we have gradience in terms of how PCWs may interact with BPT phrases in constructions where location is implicated.

Most of the time PCWs are not allowed, as in example (19a), however, example (19b) shows a situation where they are.

(19) a. **na-kàà Benísió nùùn ndasi*[4] *ñú'ùn=ra.*
IPFV-sit.down.SG Benicio face get.wet land=3SG.M
≠ 'Benicio is on the wet (area) of this land.'

b. *na-kàà Benísió nùùn vixá ñú'ùn=ra.*
IPFV-sit.down.SG Benicio face wetness land=3SG.M
≠ 'Benicio is on this land.'
= 'Benicio is on the wet part of his land.' (OR)
= 'Benicio is on the wet one of his lands (and he had multiple).'

Here, the meaning changes non-compositionally to create a partitive meaning – that is, BPT + PCW + NP triggers a possessive-like construction, probably from which this construction takes its inheritance, yielding *some* type, quantifier-like partitive meaning (i.e. either 'N$_{PART}$ *some portion of the* N$_{WHOLE}$ [+MASS] or 'N$_{PART}$ *some portion of the* N$_{WHOLE}$ [+COUNT]'). This suggests that modifying the BPT creates a new construction altogether, with the prohibition of PCW modification still intact for the more prepositional-like construction in which the BPT participates. This would constitute morphosyntactically a new form–function pairing with non-compositional meaning.

In sum, what is important here is that the PCWs may not modify BPTs in this construction in the same way as they may modify BPTs in other NP constructions.

4.8 Determiners

Determiners also provide a gradience of behaviors for BPTs used in prepositional constructions. The two main determiners in Tu'un Javi are: *meé* 'definite, specifying' and *ín* 'indefinite, singular.' In terms of referential marking, *meé* appears in definite, referential contexts. *ín* appears indefinite, in both referential and non-referential contexts. It currently seems that *meé* may not appear in non-referential contexts, but more research is needed in this regard. There is a strong preference for both determiners to come after the BPT and before its NP complement. This preference is of very high frequency. For instance, in texts, this is exclusively where determiners appear. However, the language does allow limited marking of the BPT with determiners, like it does with ordinary NPs.

[4] Again, whether attributives in Tu'un Javi constitute a distinct morphosyntactic word class apart from nouns and verbs is inchoate. This may account for the difference where a more appropriate gloss for both forms of *wet* are *vixá* 'wetness' and *ndasi* 'get wet.' Perhaps the speakers are able to more readily draw on NP+NP modifying structures rather than NP+V modifying structures, although modification by attributive Vs is certainly permitted, occurring often in discourse.

(20) a. *na-kàà* ***meé*** *tyilo'ò nùùn ñú'ù=ra*
IPFV-sit.down.SG DET boy face land=3SG.M
'The boy is on his land.'

b. *na-kàà* *tyilo'ò* ***meé*** *nùùn ñú'ù=ra*
IPFV-sit.down.SG boy DET face land=3SG.M
'The boy is ON his land (and nowhere else).'

Notice that in example (20a) the definite specifier modifies a canonical referring expression, in this case, *tyilo'ò* 'boy.' In example (20b) the definite specifier also modifies the BPT in what appears to be the prepositional construction. Here it functions to intensify the relational locative meaning of the BPT. It appears that the definite specifier functions as an additional strategy for pragmatic focus marking, as also demonstrated in examples (21a) and (21b) below.

(21) a. *kachuun Benísió nùùn ñu'ùn.*
work Benicio face land
'Benicio is on his land.'

b. *kachuun Benísió* ***meé*** *nùùn ñu'ùn.*
work Benicio DET face land
'Benicio is ON his land (and nowhere else).'

Note that the determiner *ín* is never permitted with this construction, as in example (22) below.

(22) **kachuun Benísió* ***ín*** *nùùn ñu'ùn.*
work Benicio IND face land

As expected, no *true* preposition may take determiners as shown in example (23) below.

(23) a. **kà'àn Benísió* ***meé*** *si'in Klárá*
CON.talk Benicio DET with Clara.
'Benicio is talking (very much so and only) to Clara.'

b. **kàà* *Benísió* ***meé*** *mà'ni Jwáàn tyi Júliá*
CON.sit.down.SG Benicio DET between Juan and Julia
'Benicio is sitting (very much so and only) between Juan and Julia.'

c. **kà'àn Benísió* ***ín*** *si'in Klárá*
CON.talk Benicio IND with Clara

d. **kàà* *Benísió* ***ín*** *mà'ni Jwáàn tyi Júliá*
CON.sit.down.SG Benicio IND between Juan and Julia

Here are more examples of determiner use and BPT phrases.

(24) a. *na-kàà* *Piláàr* *ini* *kárró.*
IPFV-sit.down.SG Pilar insides car
'Pilar is in her car.'

b. *na-kàà* *Piláàr* ***meé*** *ini* *kárró.*
IPFV-sit.down.SG Pilar DET insides car
'Pilar is IN her car (and nowhere else).'

c. **na-kàà* *Piláàr* ***ín*** *ini* *kárró.*
IPFV-sit.down.SG Pilar IND insides car

First, we have evidence against BPTs as canonical nouns in that they may not take the indefinite determiner. This is a feature shared with *true* prepositions. With the definite specifier *meé*, although the BPT phrase here would appear to pattern like canonical NPs, the rarity (i.e. low frequency) of its use, in my view, does not provide enough evidence to suggest anything to the contrary of BPTs in prepositional constructions. It could be that the *meé,* with its relative freedom throughout the clause, functions more like a discourse marker rather than part of the propositional grammar itself. More likely, this could simply be a *leak*, representing the gradience across word classes. While BPTs often play a role in this *other* construction (i.e. the prepositional construction), they may still retain aspects of their word class of origin. This does not, in my view, permit the denial of BPTs as common constructs within a prepositional construction. Therefore, I would consider determiner use overall as evidence, falling in favor of BPT phrases being participants in prepositional constructions.

The diagnostics reviewed here and the purpose of this section overall were to reveal certain distinct syntactic behaviors in which BPTs are implicated that differ from structures implicated in constructions of head noun/possessum 'part' and dependent noun complement/possessor (whole) relations. These diagnostics – namely, BPT strandedness, adjunction, constituency grouping, coordination, possession, quantifier modification, attributive modification, and determiner use – together demonstrate that the relationship between a specific set of entrenched and conventionalized BPTs and their NP complements is more similar to the relations between P and its complements than otherwise. However, this correlation is not without gradience in terms of potential N-schema behaviors, such as retaining the possible use of the definite specifier, in the interaction between PCWs and BPTs, and the incoherence of BPTs in a P-schema juxtaposed with a BPT in an N-schema in coordinated constructions. This *leakiness* is mostly a function of the implied historical origins of BPTs in N-schemas compared to their new emergence as a participant and constituent of P-schemas.

5. Conclusion

In this paper, I have argued that certain BPTs are implicated in the formation of prepositional constructions (or P-schema), based on the distributional behaviors which they do not share with their word class of origin (namely, nouns). These BPTs and their BPT phrases exhibit behavioral properties, not found in NP possessive constructions, including part-whole NP possessive constructions. While this paper provides substantial evidence to support the claim of BPTs as prepositional word class participants, one of the limitations of this study is the observation of BPTs in the context of multiclausal constructions. While beyond the scope of this paper, future studies

should investigate this area of inquiry. Parallel behavioral innovations among a semantic natural class of lexemes are what constitutes this prepositional word class. This process is inherently fuzzy because it is dynamic and emergent. Lexemes are categorically unspecified elements which participate in schematic constructions at such frequencies that the strength of associations leads to syntactic reanalysis and grammaticalization. This *leakiness* is mostly a function of the implied historical origins of BPTs in N-schemas compared to their new emergence as participants and constituents of P-schemas.

References

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Jacobs, Roderick & Rosenbaum, Peter (eds.), *Readings in English Transformational Grammar*, 184-221. Waltham: Ginn.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Deacon, Robert Joel. 2014. Adpositions and lexical categories: Distributed Morphology's insight into the problem. *Florida Linguistics Papers* 1. 1-12.

Delancey, Scott. 2005. Adpositions as a non-universal category. In Frajzyngier, Zygmunt & Hodges, Adam & Rood, David S. (eds.), *Linguistic diversity and language theories*, 185-202. Amsterdam: John Benjamins.

Diessel, Holger. 2020. A dynamic network approach to the study of syntax. *Frontiers in Psychology* 11(604853). 1-15.

Dryer, Matthew S. 2007. Clause types. In Shopen, Timothy (ed.), *Language typology and syntactic description,* vol. 1, 2nd edn, 224-275. Cambridge: Cambridge University Press.

Evans, Nicholas & Osada, Toshiki. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9(3). 351-390.

Frajzyngier, Zygmunt. 2022. Toward a functional typology of adpositions: Theoretical implications. *Folia Linguistica* 56(1). 123-151.

Givón, Thomas. 2021. *The life cycle of adpositions*. Amsterdam: John Benjamins.

Goldberg, Adelle. 1995. *A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adelle. 2006. *Constructions at work*. Oxford: Oxford University Press.

Grimshaw, Jane. 1991. *Extended projection*. Ms. Brandeis University.

Grimshaw, Jane. 2005. *Words and structure*. Stanford: CSLI Publications.

Hagège, Claude. 2010. *Adpositions*. Oxford: Oxford University Press.

Hollenback, Barbara E. 1995. Semantic and syntactic extensions of body-part terms in Mixtecan: The case of "face" and "foot." *International Journal of American Linguistics* 61(2). 168-190.

Jacobsen, William H. 1979. Noun and verb in Nootkan. In Efrat, Barbara S. (ed.), *The Victorian conference on Northwestern languages*, 83-153. Victoria: British Columbia Provincial Museum.

Lillehaugen, Brook Danielle. 2003. Syntactic and semantic development of body part prepositions in Valley Zapotec languages. In Castillo, Jeanie (ed.), *Santa Barbara papers in linguistics: Proceedings from the sixth workshop on American Indigenous languages*, vol. 14, 69-82. Santa Barbara: University of California at Santa Barbara.

Lillehaugen, Brook Danielle. 2006. *Expressing location in Tlacolula Valley Zapotec*. University of California at Los Angeles. (Doctoral dissertation.)

Lillehaugen, Brook Danielle & Sonnenschein, Arron Huey. 2012. Expressing location in Zapotec: An introduction. In Lillehaugen, Brook Danielle & Sonnensche, Aaron Huey (eds.), *Expressing location in Zapotec*, 1-34. Munich: Lincom.

Munro, Pamela. 2006. From parts of speech to the grammar. In Payne, Thomas E. & Weber, David J. (eds.), *Perspectives on grammar writing*, 71-111. Amsterdam: John Benjamins.

Penner, Kevin L. 2020. *Adjectives in Ixtayutla Mixtec*. Ms. SSILA.

Rijkhoff, Arrow Jan & Van Lier, Eva. (eds.). 2013. *Flexible word classes: Typological studies of underspecified parts of speech*. Oxford: Oxford University Press.

Sapir, Edward. 1921. *Language*. New York: Harcourt Brace.

Schmid, Hans-Jörg. 2015. A blueprint of the entrenchment-and-conventionalization model. *German Cognitive Linguistic Association* 3. 3-25.

Thompson, Sandra A. 1988. A discourse approach to the cross-linguistic category 'adjective.' In Hawkins, John A. (ed.), *Explaining language universals*, 167-185. Oxford: Blackwell.

Vandeloise, Claude. 1991. *Spatial propositions: A case study from French*. Chicago: University of Chicago Press.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:2.51-64 (2025)**

LACUS Forum is published on at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Morphological complexities and sociolinguistic barriers against the adoption of feminatives in contemporary Polish public discourse from the Gen Z perspective

Artur Urbaniak
Adam Mickiewicz University

Abstract: The Polish language in public discourse and politics is currently undergoing a transformation with the introduction, or rather a reintroduction, of feminatives (i.e. forms that denote feminine noun forms for job titles, social roles and functions). In contrast to English, which trends toward gradual gender-neutralization, Polish adopts strong gender distinctions to emphasize the role and significance of women in society, specifically in politics, public offices, and academia. This shift has sparked controversy, resulting in a polarization between supporters and opponents of the use of feminatives in formal contexts. This paper aims to demonstrate that resistance to these newly created forms may stem not solely from political or ideological motives, but may also be influenced by practical, linguistic considerations. The survey results indicate that the current generation of 20-year-olds (Gen Z) – though they generally express a willingness to use feminatives, particularly when there are equivalents in the language that appear linguistically acceptable in their form – might actually refrain from using them due to purely pragmatic considerations.

Keywords: sociolinguistics, feminatives, inclusive language, language change, feminization, pragmatics

Languages: English, Polish

1. Introduction

WITH ENGLISH INCREASINGLY ADOPTING predominantly gender-neutral noun forms (Grigoryeva & Nazmieva 2023), the linguistic processes currently unfolding in Polish present a compelling contrast. While English trends toward gender neutrality, eliminating its limited gender-specific noun forms (such as *postman* or *chairman* becoming *postal worker* and *chair*), Polish is moving in the opposite direction, with a growing emphasis on nouns explicitly designating gender – specifically with female referents – through distinct morphological gender markers. In feminist theory, language is often viewed as a powerful tool of patriarchy (Pauwels 1998, 1999), and this perspective has spurred the reintroduction of certain linguistic forms to enhance the visibility of women in society. For example, the term *herstory* has been coined as a counterpart to *history*, underscoring that history is not solely about men (Pauwels 1998). This broader shift toward

gender-neutrality in English manifests in terms such as *actress*, *waitress*, or *stewardess* being replaced by neutral equivalents like *actor*, *waiter* (or *server*), and *flight attendant*. The same trend applies to traditionally masculine terms, with *fireman*, *policeman*, and *postman* giving way to gender-neutral terms like *firefighter*, *police officer*, and *mail carrier* (cf. Wtorkowska 2019, 2020; Linde-Usiekniewicz 2022).

In contrast, the Polish language maintains a complex grammatical gender system, encompassing masculine-personal and non-masculine-personal categories in the plural, as well as masculine, feminine, and neuter in the singular. This division originates from historical morphological developments and remains a defining feature of its grammar. Polish grammatical structure links gender closely with inflectional endings, differentiating gender in the plural through masculine-personal and non-masculine forms, based on the presence of male individuals in a group. This aspect influences the forms of verbs, adjectives, and numerals, setting Polish apart from many other Indo-European languages that lack this level of gender marking (Długosz-Kurczabowa & Dubisz 2003). The first tendency, the apparent drive in English to equalize grammatical forms so that they are gender-neutral, is referred to as *neutralization*. The second trend, evident in Polish, towards forming gender-specific nouns, is referred to as *feminization* (Karwatowska & Szpyra-Kozłowska 2005; Szpyra-Kozłowska 2019).

Contemporary English largely lacks grammatical gender, with distinctions primarily limited to biological sex in pronouns (e.g. *he, she, they*). Historically, English, like other Germanic languages, once used three grammatical genders – masculine, feminine, and neuter – that influenced noun and adjective forms. Over time, however, English underwent simplification, abandoning grammatical gender and adopting a streamlined structure. Today, gender in English is reflected only through personal pronouns and aligns more with biological sex than with grammatical gender (Corbett 1991; Smith 2023). In contrast to English, which lacks gender marking within nouns and adjectives, Polish grammar makes gender a core component, necessitating both feminine and masculine forms for nouns that denote people. This distinction is especially relevant in discussions of feminatives, as Polish explicitly requires grammatical gender for each personal noun. For years the Polish language has adopted the tradition of using masculine-personal forms to refer to people in general, regardless of gender. Hence, it is correct to use a sentence: *Anna jest lekarzem* 'Anne is a doctor' from *lekarz* 'a male doctor,' even though there is a feminine-personal form (feminative) *lekarka,* and one can say *Anna jest lekarką* 'Anne is a female doctor' (cf. Łaziński 2006; Stępień 2018).

Accordingly, Polish has long included feminine forms for a wide range of professions and social roles held by women, such as *lekarka* 'female doctor,' *nauczycielka* 'female teacher,' and *sekretarka* 'female assistant.' However, certain high-prestige titles, professions, and social roles in politics and academia still lack widely accepted feminine equivalents, with male forms historically serving as the default for individuals of both genders. This study examines Generation Z's perceptions and offers insights into how feminatives are received among younger Polish speakers (cf. Dubisz 2021; Szpyra-Kozłowska 2021; Linde-Usiekniewicz 2022).

In the face of the heated debate surrounding the introduction or elimination of feminatives from language, it is worth noting what the findings suggest that Generation Z's attitudes toward feminatives are shaped not only by political or ideological divides – factors more prevalent among earlier generations – but also by morphological challenges. Complex word-formation processes and the absence of standardized patterns for creating feminatives (e.g. *adjunktka* for 'female adjunct' or *architektka* for 'female architect') create linguistic/morphological obstacles, potentially leading younger speakers to avoid feminatives less out of perceived irrelevance and

more due to uncertainty about appropriate or normative forms.

2. Understanding feminatives in Polish: Definition and historical development

The term *feminatives* (Polish *feminatywy*, derived from Latin *feminativum/-a*, meaning 'made effeminate,' or *feminativus*, meaning 'making effeminate') is part of specialized linguistic vocabulary (Dubisz 2021: 111). In Polish, it refers to nouns typically derived from masculine forms that denote a female profession, role, or function or serve as the feminine equivalent of terms used to describe agents of actions or specific personal roles associated with women (cf. Łaziński 2006). Feminatives include nouns that mark female-specific identities in roles, professions, or societal functions, usually through suffixes that align with the gendered structure of Polish grammar (Dubisz 2021: 112; cf. Szpyra-Kozłowska & Laidler 2022; Łaziński 2023). Common suffixes used to form feminatives include *-ka, -owa,* and *-yni,* as seen in *kasjer – kasjerka* 'male/female cashier,' *księgowy – księgowa* 'male/female accountant,' and *sprzedawca – sprzedawczyni* 'male/female shop assistant' (see Małocha-Krupa 2021: 103; Linde-Usiekniewicz 2022: 82).

Historically, feminatives have been part of the Polish language since at least the 16th century, supported by the flexibility of the Slavic morphological system, which inherently allows for the creation of gender-specific forms. This system's capacity to generate masculine, feminine, and neuter nouns provided a structural foundation for constructing feminatives through gendered suffixes, making it possible to produce distinct forms for various genders (Dubisz 2021). One of the earliest records of feminatives appears in *Dictionarius Ioannis Murmelli* (1526), which includes terms like *tkaczka* 'female weaver' and *włodarka* 'female head of the household,' showing that feminine forms were linguistically and socially relevant in Old Polish (Małocha-Krupa 2018: 18; Jankowski 2021: 74). By the early modern period, feminatives had developed into a distinct set within the Polish lexicon, regularly using suffixes like *-ka* and *-owa* but also *-yni, -yca* and *-ini* to denote female-specific roles (Dubisz 2021; Rodek 2023). Some feminative terms date back as early as the sixteenth century, evidenced in recorded terms like *złotniczka* 'female goldsmith' and *stolarka* 'female carpenter' (Osińska-Szymańska & Preuhs 2022: 12).

The 20th century also saw instances of specific feminatives emerging, such as *doktorka* 'female doctor,' which was mentioned as early as 1901 in *Poradnik Językowy*. This form highlighted the growing need for feminine linguistic equivalents as women entered professional fields traditionally dominated by men (Lisczyk 2023). Feminatives have thus evolved alongside social movements, facing both advocacy and resistance in public and academic discourse. During the interwar period, feminatives gained wider usage, reflecting broader European feminist movements advocating for the linguistic visibility of women in professional and public spheres. It was during this time that terms like *adwokatka* 'female lawyer' began to emerge, marking evolving social roles for women and a growing recognition of the need for gender-specific linguistic representation (cf. Sieczkowski 1934; Małocha-Krupa 2021). However, during the communist period in Poland, feminative usage declined, as state policies favored gender-neutral forms to reflect socialist ideals of gender equality, with feminatives often relegated to lower-status occupations (Brzyszcz 2019; Satoła-Staśkowiak & Sosnowski 2019; Jankowski 2021; Szpyra-Kozłowska 2021; Szpyra-Kozłowska & Laidler 2022).

The historical trajectory of feminatives in Polish reveals a broader narrative of social and ideological tensions, reflecting shifting gender roles and values within society. From their use in early Polish, through the restrictions of the Polish People's Republic era, to current debates over

their relevance and legitimacy, feminatives represent an ongoing conversation between language and social identity. Their persistence – and the debates they provoke – indicates that feminatives may continue to grow as part of Polish, influenced by ongoing progress toward gender equality in society and language alike.

3. Opinions of the Council for the Polish Language and social acceptance of feminative forms

The Council for the Polish Language (*Rada Języka Polskiego*), an opinion and advisory body on the use of the Polish language which renders opinions on various linguistic forms and language use, has issued two significant positions regarding feminatives, reflecting evolving perspectives on their usage within Polish. These stances, taken in 2012 and 2019, highlight a shift influenced by social, political, and linguistic factors.

The initial stance from March 2012 expressed a conservative view on feminatives. This position discouraged the widespread use of feminative forms, arguing that many new feminatives felt linguistically awkward, lacked standardization, or carried colloquial or diminutive connotations that made them less suitable for official and formal contexts. For instance, while forms like *psycholożka* 'female psychologist' or *architektka* 'female architect' technically fit within Polish word-formation rules, the Council indicated that such forms might seem overly informal or even trivializing in certain professional contexts.[1] Thus, the Council initially advised limiting feminatives to less formal contexts, emphasizing that masculine forms could serve as gender-neutral terms in formal and professional settings (cf. Latos 2019; Skudrzyk 2020).

By contrast, the November 2019 statement marked a departure from the restrictive tone of the earlier stance. This revised position acknowledged an increased public and academic interest in linguistic gender equality and proposed a more inclusive approach. Recognizing the cultural and ideological factors driving language evolution, the Council suggested that feminatives could serve an important social role by enhancing gender visibility.[2] The Council recommended promoting both feminine and masculine forms in parallel, allowing for symmetrical gender representation in the language without mandating uniform changes across all contexts (cf. Brzyszcz 2019; Latos 2019; Jarosz 2021). The Council's 2019 opinion also emphasized the flexibility of the language, stressing that choice should be guided by user preferences and contextual needs rather than strict linguistic norms. This approach allowed for both established and emerging forms, balancing tradition with modern social values. Overall, the Council's evolving positions reflect the complexities of Polish linguistic norms, shaped as much by cultural and social changes – such as society's shift toward greater democratization and liberalization – as by structural linguistic considerations (Latos 2020).

Linguistic change often prompts reactions of either acceptance or rejection from language users. It is not possible to impose specific language forms on society; rather, these forms must emerge gradually through education and by raising awareness among language users, as part of the natural process of language evolution. This is because language inherently reflects the speaker's perception of reality. As a consequence, Polish society remains polarized on the issue of feminative forms, with resistance to newly formed feminatives partly due to linguistic conservatism and

[1] Position of the Council of the Polish Language on the feminine forms of the names of professions and titles adopted at the plenary meeting of the Council on March 19, 2012.

[2] Position of the Council of the Polish Language on the feminine forms of the names of professions and titles adopted at the Council's plenary meeting on November 25, 2019.

political tensions around gender. Terms like *marszałkini* 'female marshal' or *ministra* 'female minister' still face opposition from more conservative language users, although younger generations are generally more accepting (cf. Skudrzyk 2020; Łaziński 2023: 348). Traditionalists often dismiss feminatives as artificial or ideologically motivated, while proponents argue they are essential for linguistic equality. Surveys show divided public opinion, especially around terms like *profesorka* 'female professor' or *gościni* 'female guest,' with younger generations showing greater acceptance (Małocha-Krupa 2018; Szpyra-Kozłowska 2021).

Notably, social acceptance of feminatives is split, with some viewing them as necessary for gender equality, while others resist them as artificial or ideologically driven, perceiving language as a powerful tool of mind control (McConnell & Fazio 1996). Social media has intensified public debate, with feminist groups advocating for feminatives as a component of inclusive language practices (Brzyszcz 2019: 86; Jarosz 2021: 996). Moreover, beyond political divisions, feminatives often encounter social resistance, with some considering them informal or humorous and some perceiving them as diminutives, for example, *ministerka* 'female minister,' due to the use of the suffix *-ka*, typically used to create diminutives (as in *śruba – śrubka* 'a screw/ a small screw'; *rana – ranka* 'a wound/ a small wound'). This view, while not fully morphologically justified, seems understandable from the standpoint of an average language user.

Synthesizing these perspectives, it is clear that this polarizing perception, often reflected in media discussions, frames feminatives as either linguistic progress or unnecessary deviations from the norm (cf. Stępień 2018; Brzyszcz 2019; Szpyra-Kozłowska 2021). Regardless of the opinion upheld by the Council for the Polish Language, a rational approach leads to the conclusion that young, educated people holding more liberal views on language use, and adopting more inclusive social approach, would strongly advocate for the use of progressive and egalitarian feminine forms.

4. Study design, findings, and analysis

The aim of this study was to test the hypothesis that young, educated members of Generation Z (1) demonstrate a stronger preference for inclusive language, reflected in their openness to using feminatives, and (2) lack formal knowledge of the rules for creating feminatives in their native language, often relying instead on linguistic intuition.

To test this hypothesis, an anonymous online survey was conducted with 147 students from two universities in Poland, gathering responses on their attitudes toward feminatives. The survey included both multiple-choice and open-ended questions, asking respondents to create or suggest feminine forms from masculine forms of selected nouns representing social or professional roles, or functions. The following sections present a detailed analysis of the survey results.

4.1 Demographic profile of participants

The survey was administered to a sample of 147 students (n=147) across two urban centers: one with a population exceeding 600000 residents, and the other with a population of fewer than 100000 residents. Among the respondents, a majority identified as female (71.9%), while males constituted a minority (23.3%), and 4.8% of participants chose not to disclose their gender. Respondents were representative of diverse population backgrounds across various settlement sizes: 26.5% were from large cities (over 500000 residents), 21.8% from small towns (1000 to 10000 residents), 19.7% from medium-sized towns (50000 to 100000 residents), 17% from smaller urban centers (10000 to 50000 residents), 8.8% from rural villages (up to 1000 residents), and the smallest group (6.2%) represented medium-sized urban centers (100000 to 500000 residents).

This distribution indicates a demographically varied sample with respect to gender and geographic origin. Age-wise, a significant majority of participants (70.8%) were between 20 and 25 years, with the remaining 29.2% being under the age of 20.

4.2 Employment and education status

In terms of current work and study situations, 49% of respondents reported they were exclusively studying, 41.5% combined work with study, and 9.5% were actively studying while also seeking employment.

4.3 Language Background

Polish was the native language for a substantial majority of participants (91.2%). Among the remainder, 4.7% identified Ukrainian as their first language, 1.4% identified Russian, and 0.7% reported Italian, and another 0.7%, German. Additionally, 1.4% of respondents did not specify their native language, but indicated that it was not Polish. The participant demographics demonstrate a diverse sample across gender, regional origin, age, and linguistic background, supporting a representative basis for analyzing attitudes within this population.

4.4 Study findings: Academic titles, functions and forms of address at universities

In Polish academic and public spheres, the question of formal address in university contexts, particularly regarding how to address female professors, has become a topic of significant discussion. This debate reflects wider socio-political sensitivities, as academic circles remain divided between traditional and feminative forms.

Survey respondents were asked to identify the preferred form of address for female professors. The responses indicated a strong preference for the traditional form *Pani Profesor* 'Ms Professor,' which garnered 67.3% of the responses. Feminative forms such as *Profesorka* and *Profesora* received less support, with 18.4% and 9.5%, respectively. The feminative forms paired with the polite term *Pani*, such as *Pani Profesora* and *Pani Profesorka*, received minimal support (3.4% and 0.7%, respectively). Only one response favored the masculine form *Profesor* (0.7%), as shown in Table 1.

Table 1: Preferred female professors' titles in the university context

Form of address	%
Pani Profesor	67.3
Profesorka	18.4
Profesora	9.5
Pani Profesora	3.4
Pani Profesorka	0.7
Profesor	0.7

These findings underscore the existing divide within the academic and public spheres regarding gendered titles, with the majority still gravitating toward established forms that do not employ feminative endings.

Similarly, when asked about the preferred form of address for women with a doctoral degree, the traditional term *Pani Doktor* was again the most popular choice, with 67.3% of respondents

selecting it. Feminative forms *Doktora* and *Pani Doktora* were chosen by 11.6% and 7.5% of participants, respectively. Interestingly, the feminative *Doktorka* and its variant *Pani Doktorka* were less favored, receiving only 5.4% and 0.7% of responses, placing them below the masculine form *Doktor* (7.5%). Detailed responses are shown in Table 2.

Table 2: Preferred female doctoral titles in the university context

Form of address	%
Pani Doktor	67.3
Doktora	11.6
Pani Doktora	7.5
Doktor	7.5
Doktorka	5.4
Pani Doktorka	0.7

These findings reveal a strong preference among respondents for traditional forms of address such as *Pani Profesor* and *Pani Doktor*. Fewer than one-third of respondents expressed willingness to use feminative forms.

When addressing senior university positions in the Polish university setting, established social norms strongly influence language use, particularly in addressing faculty members and administrative officials in formal correspondence or interactions. Traditionally, male-gendered forms have been used for academic titles and roles, such as *Pani Profesor* for female professors and *Pani Doktor* for female doctors, with the feminine designation achieved through the addition of *Pani* to the masculine noun. This tradition also extends to higher-ranking university titles, such as *Pani Dziekan* 'Ms Dean' or *JM Rektor* 'Her Magnificence Rector' for a female rector. In Polish, *JM* can stand for either *Jej Magnificencja* 'Her Magnificence' or *Jego Magnificencja* 'His Magnificence,' followed by the title *Rektor* 'Rector' or its feminine forms *Rektora* or *Rektorka* 'female Rector.'

However, recent years have seen a gradual shift in these conventions, with feminative forms emerging, including *profesora*, *docentka*, *doktora*, and *dziekana*, indicating a move toward more gender-specific language for these roles.

For the title of *JM Rektor* held by a woman, respondents were divided. While 30.1% supported the traditional form *JM Rektor*, 24.7% preferred the variant *JM Pani Rektor*. Notably, 24.7 chose the feminative form *JM Rektorka*, and 19.8 selected *JM Rektora*. Overall, 54.8% favored traditional forms, either *JM Rektor* or *JM Pani Rektor*, while 44.5% preferred feminative forms *JM Rektorka* or *JM Rektora*. These results are presented in Table 3.

Table 3: Preferred female Rector titles in the university context

Form of address	%
JM Rektor	30.1
JM Pani Rektor	24.7
JM Rektorka	24.7
JM Rektora	19.8

This distribution suggests that while traditional forms remain slightly more popular, a significant portion of respondents is open to feminative alternatives, reflecting ongoing linguistic and social shifts in Polish academic norms.

4.5 Study findings: Forms of address in political contexts

In contrast to academic settings, respondents displayed a stronger preference for feminative forms when referring to women in political roles. When asked to select the appropriate term for a female politician, participants were offered both the traditionally masculine form *polityk* 'male politician' and the feminative options *polityczka* and *politykini* (using feminine endings -a and -ini, respectively). Additionally, the combination *Pani polityk* 'Ms politician' was available as a gender-indicating, yet traditionally masculine, option.

Analysis reveals that feminative forms (*polityczka* 51% and *politykini* 2.7%) collectively received the majority (53.7%) of responses, surpassing the masculine form *polityk* (43.5%) (Table 4). This suggests a greater acceptance of feminative terms in political contexts than in academic ones, where respondents had previously favored more traditional forms.

Table 4: A term referring to a female politician

Form of address	%
Polityczka	51
Polityk	43.5
Politykini	2.7
Pani polityk	2.1

A similar trend emerged regarding the position of *minister* in the government. Respondents were asked to indicate the most suitable term for a female minister. Here, 42.9% selected *ministra* (a feminative form with the feminine *-a* ending), and 10.9% chose *ministerka* (another feminative with the -ka ending), resulting in a combined 53.8% preference for feminative forms. Meanwhile, 46.3% opted for traditional forms, with 35.4% selecting *Pani minister* and 10.9% choosing the masculine *minister*. Although feminative forms were slightly more popular, the distribution suggests a nearly equal split (Table 5).

Table 5: A term referring to a female minister

Form of address	%
Ministra	42.9
Pani minister	35.4
Minister	10.9
Ministerka	10.9

Table 6: A term referring to a female Speaker of the House

Form of address	%
Pani Marszałek	38.8
Marszałkini	38.8
Marszałek	10.9
**Marszałka*	5.8
**Pani Marszałka*	2.1
Pani Marszałkini	0.7

For the role of *Marszałek Sejmu* 'Speaker of the Sejm' (the lower house of the Polish parliament), respondents were more evenly divided. Traditional forms, such as *Pani Marszałek*

(38.8%) and *Marszałek* (10.9%), received considerable support. However, feminative options, including *Marszałkini* (38.8%) and *Marszałka* (5.8%), as well as hybrid forms like *Pani Marszałka* (2.1%) and *Pani Marszałkini* (0.7%), collectively accounted for nearly half of the responses, showing a strong openness to feminative alternatives (see Table 6).

When participants were asked to suggest the appropriate feminative form for *burmistrz* 'male mayor,' another position in public administration, their responses were as in Table 7:

Table 7: A term referring to a female mayor

Form of address	%
Burmistrzyni	49.3
Pani burmistrz	38.2
Pani burmistrz/burmistrzyni	2.8
Other forms	4.2

These findings show that for *burmistrz*, a majority (49.3%) preferred *burmistrzyni* as the feminative form, while 38.2% selected the gender-neutral *Pani burmistrz*. This distribution suggests that while there is strong support for feminative forms in political and public administration contexts, traditional and hybrid forms still hold significant ground.

4.6 Study findings: Attitudes towards feminatives

In response to the question: *Are you in favor of using feminine personal forms (so-called feminatives) in the public sphere (e.g. in higher education institutions, government offices, seats of power, politics)?*, over 53% of respondents answered *yes, but only if the feminine form sounds reasonable and is already established in the language*. Another 21% responded with *definitely yes*, while 17% chose *rather yes*. This reflects a linguistic pragmatism among Polish speakers, as there is a prevailing tendency to use feminine forms when they are phonologically and grammatically straightforward to create. Only slightly over 6% of respondents expressed opposition to the use of feminine forms, and 2% were undecided on the issue.

In a concluding question, respondents were asked to express their position on feminatives by selecting one or more of the available statements. The responses are presented in Table 8.

These responses reveal a nuanced range of opinions among young Poles regarding feminatives, with a pragmatic approach prevailing: support for feminatives exists when they are perceived as phonologically accessible, grammatically compatible, and socially familiar. Resistance is generally tied to perceptions of impracticality, humor, or ideological motives. The data reflects both support for gender visibility and a desire for linguistic consistency within Polish.

Table 8: Attitudes toward the adoption of feminatives

Suggested Opinion	%
I believe feminatives are an important issue in contemporary Polish, as they highlight the growing importance of women in the broader public sphere.	39.5
I make an effort to always use the appropriate feminine forms when referring to women in my statements; for example, *my mother is a psychologist* (*psycholożka*), *my sister is a scientist* (*naukowczyni*).	37.4
Some words, such as *chirurżka* 'female surgeon,' *adiunktka* 'female adjunct,' or *architektka* 'female architect,' seem difficult to pronounce and impractical.	40.1
In titles and positions, I prefer using the form *Pani + title or position*, e.g. *Pani Profesor* 'Ms. Professor,' *Pani Marszałek* 'Ms. Marshal,' or *Pani Burmistrz* 'Ms. Mayor.'	53.7
Gender should definitely be reflected in titles, e.g. *JM Rektor Jan Nowak*, but *JM Rektora Janina Nowak*.	16.3
I use feminine forms only when they have an established tradition, such as *lekarka* 'female doctor,' *pielęgniarka* 'nurse,' *asystentka* 'assistant,' *nauczycielka* 'teacher,' but not in cases like *archeolożka* 'female archaeologist' or *pedagożka* 'female pedagogue.'	29.3
I do not object to the use of feminatives, but I do not always use them personally because I am unsure of the correct term, e.g. *kierowca autobusu* 'bus driver,' but should a woman be called *kierowczyni*?	39.5
I think that titles using *Pani + title* sound more professional than titles with feminatives; for example, I prefer *Pani Minister* 'Ms. Minister' over *Ministra* or *Ministerka*.	29.3
I believe the issue of feminatives is an artificial problem and has no impact on reality.	12.2
I never use feminine forms for professions and roles traditionally held by men, such as *mechanik* 'mechanic' or *hydraulik* 'plumber.'	6.8
I believe that the masculine personal noun in Polish, e.g. *malarz* 'painter,' *piekarz* 'baker,' or *mechanik* 'mechanic,' refers to a person performing this job regardless of gender.	24.5
I consider some feminine forms to be strange or humorous, such as *adiunktka* 'female adjunct' or *ministra* 'female minister.'	27.9
I have no opinion.	2.1

4.7 Discussion of the results

A detailed analysis of the findings leads to several insights reflecting broader attitudes of Generation Z toward the use of feminatives in public discourse. The first observation reveals a slight discrepancy between the declared willingness to use feminatives and the practical application of these forms. On a declarative level, over 90% of respondents express a willingness to use feminatives (21% are decidedly in favor, 53% are open to usage provided the form is phonologically and grammatically acceptable, or simply usable, and 17% are somewhat willing). Only 6% of respondents report no desire or need to use feminatives, while 2% are undecided. However, when these responses are compared with data on the actual usage of feminatives in academic contexts, a preference for traditionally masculine forms, prefaced with *Pani* 'Ms.,' is evident. For instance, *Pani doktor*/ *Doktor* (67% and 7%, respectively), *Pani Profesor* (67%), *JM Rektor*/ *JM Pani Rektor* (30% and 16%, respectively) are preferred.

These results should be interpreted within the context of traditional forms, noting that not all universities have adopted a standardized system for addressing women in academia or in official roles. Consequently, students may simply be uncertain about which feminative is correct (e.g. *Profesora* vs. *Profesorka* or *Doktora* vs. *Doktorka*). As a result, they tend to adhere to established titles. Moreover, academic women themselves may have specific preferences for how feminatives are used. Interestingly, more than 53% of respondents report a preference for using forms with the prefix *Pani* in formal titles, such as *Pani Marszałek* or *Pani Minister*.

In the political discourse, regarding government or parliamentary positions, deeply ingrained divisions are observed. For instance, respondents express a preference for the feminative *burmistrzyni* (49%) over *Pani burmistrz* (38%), demonstrating linguistic creativity by also suggesting novel forms such as *burmistrza*, *burmistrka*, or *burmistrzka* – terms not found in the Polish dictionary. This reflects a readiness to use feminatives in public spaces, even if there is some uncertainty about their formation. Similarly, for the position of *minister*, the feminative *ministra* (43%) is preferred over *Pani minister* (35%), with *ministerka* and the masculine form *minister* each receiving 10%. However, in the case of the *marszałek* of the Sejm, respondents are more divided, with the masculine form *Pani Marszałek* (39.8%) and the feminine form *Marszałkini* (39.8%) receiving equal support, while 11% favor the masculine form *Marszałek* without the prefix *Pani*.

Approximately 40% of respondents consider feminatives important as they reflect the role of women in society and public domains (e.g. academia and politics). Over 37% value the use of feminatives in spoken language, seeing it as a vital component of linguistic world-creation (e.g. *Moja mama jest psycholożką* 'My mom is a female psychologist' instead of *Moja mama jest psychologiem* 'My mom is a psychologist,' using the feminative *psycholożka* derived from the masculine *psycholog*). Nearly 28% of respondents find some feminatives strange or unnatural-sounding, which reflects the role of linguistic habits in the perception of new word forms. Forms that have already permeated social awareness tend to be more widely accepted.

A pragmatic approach emerges among respondents, who report that some feminatives sound awkward and seem impractical or challenging to pronounce. Close to 40% share this view, providing examples like *chirurżka*, *adiunktka*, or *architektka*. About 40% also report that, while they do not oppose the use of feminatives, they are not always ready to use them due to uncertainty about the correct form (e.g. for *kierowca*, they are unsure of an appropriate feminine version like *kierowczyni*). At the same time, nearly one-third of respondents state that they use only well-established feminatives, such as *nauczycielka*, *pielęgniarka*, or *lekarka*, but avoid others like *psycholożka* or *archeolożka*.

These findings indicate that while there is a general openness toward feminatives among Generation Z, actual usage is influenced by factors such as tradition, familiarity, and individual preferences. Feminatives with established roots in the language gain more acceptance, whereas newer forms often encounter resistance, sounding awkward or unnatural. Additionally, the inclination to use feminatives is context-dependent, showing variance across domains like academia and politics. This highlights an overarching trend: while the younger generation shows a readiness to integrate feminatives into the public sphere, the process is gradual, with acceptance relying on linguistic standardization and social exposure. As feminatives become more common in media and daily discourse, there is potential for broader societal acceptance, especially if linguistic preferences align with ease of use and cultural resonance.

4.8 Study limitations

The present study, conducted across various academic centers with a sample of 147 participants, provides valuable insights but does not permit the drawing of broad, generalizable conclusions. Future research should replicate this study with a larger and more diverse sample to enhance its validity. A notable limitation is the demographic composition of the sample, as over 70% of respondents were women. Furthermore, the study was conducted among representatives of Generation Z within academic settings, predominantly involving individuals who are either educated or currently pursuing higher education. As such, the findings should be interpreted with these contextual limitations in mind.

5. Conclusion

The study aimed to assess Generation Z's attitudes toward feminatives in Polish public discourse, positing that younger Poles are open to using feminatives but may lack knowledge about their formation. Survey results indicate a notable openness toward feminatives; however, respondents exhibit a discrepancy between their declared intent and actual usage. While over 90% express willingness to use feminatives, their usage often reverts to traditional masculine forms prefaced by *Pani* (e.g. *Pani Profesor*, *Pani Doktor*), particularly in academic contexts. This suggests an adherence to established forms due to familiarity or uncertainty regarding correct feminative forms.

Respondents show more linguistic creativity in political discourse, favoring feminative forms like *burmistrzyni* for mayor over the traditional *Pani burmistrz*. This indicates an openness to using feminatives in public roles, albeit with variability in preferred forms. The preference for *ministra* (43%) over *Pani minister* (35%) and equal support for *Marszałkini* (feminine 'Speaker of the House') alongside *Pani Marszałek* highlight a readiness to adopt feminatives, even if respondents are unsure of established forms.

Close to 40% of respondents consider feminatives important, viewing them as reflections of women's roles in society. However, a similar portion finds feminatives awkward or impractical, and many report difficulty with specific feminatives like *chirurżka* ('female surgeon') or *architektka* ('female architect'). Respondents lean toward feminatives with strong traditional roots (e.g. *nauczycielka*, *pielęgniarka*) over newer, less established forms.

This study reveals that Generation Z displays a general openness to feminatives, shaped by pragmatism and linguistic familiarity. Usage of feminatives varies by context, with greater acceptance in political discourse than in academic settings. Feminatives that are linguistically intuitive or already established enjoy broader acceptance, while newer or phonologically

challenging forms encounter resistance. The findings suggest a gradual societal shift toward feminative usage, contingent on continued linguistic standardization and exposure. Further research is warranted to explore the sociolinguistic factors influencing the evolving acceptance of feminatives in Polish.

References

Brzyszcz, Anita. 2019. Kreatywność leksykalna w zakresie tworzenia form feminatywnych przez użytkowników portalu Facebook komentujących najpopularniejsze feministyczne fanpage'e [Lexical creativity in the creation of feminative forms by Facebook users commenting on the most popular feminist fanpages]. In Burska, Katarzyna & Cieśla, Bartłomiej (eds.), *Kreatywność językowa w komunikacji internetowej* [Linguistic creativity in online communication], 85-98. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Corbett, Greville G. 1991. *Gender*. Cambridge: CUP.

Dubisz, Stanisław. 2021. Raz jeszcze o feminatywach i feminizmie – uwagi spokojne [Once again about *feminatywy* 'feminativa' and *feminizm* 'feminism': calm comments]. *Poradnik Językowy* 2021(5). 111-113.

Grigoryeva, Leona & Nazmieva, Elmira. 2023. The gender aspect of the English language. Should we teach gender neutrality in the classroom? *Journal of Education Culture and Society* 14(2). 144-153. https://doi.org/10.15503/jecs2023.2.144.153

Jankowski, Hubert. 2021. W obronie języka czy w obawie przed zmianą? Manipulacje językowe w prawicowych tekstach publicystycznych o nazwach żeńskich [To protect the language or out of the fear of change? Language manipulations in right-winged journalistic texts about feminine nouns]. *Acta Universitatis Lodziensis. Folia Linguistica* 55. 73-87. https://doi.org/10.18778/0208-6077.55.04

Jarosz, Marta. 2021. Działania aktywistów języka inkluzywnego w mediach społecznościowych a społeczne postrzeganie feminatywów [Activities of inclusive language activists in social media and the social perception of feminatives]. *Studia medioznawcze* 2(3). 995-1015.

Karwatowska, Małgorzta & Szpyra-Kozłowska, Jolanta. 2010. *Lingwistyka płci. Ona i on w języku polskim* [*Gender linguistics. She and he in the Polish language*]. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Latos, Agnieszka. 2020. Feminatywy w stanowiskach Rady Języka Polskiego. Język a ewolucja normy społecznej [Feminatives in the positions of the Council of the Polish language. Language and the evolution of the social norm]. *Postscriptum Polonistyczne* 26(2). 227-242. https://doi.org/10.31261/PS_P.2020.26.18

Linde-Usiekniewicz, Jadwiga. 2022. The case for feminatives in Polish. *ACADEMIA. The magazine of the Polish Academy of Sciences* 4(76). 82-82. https://doi.org/10.24425/academiaPAS.2022.144690

Łaziński, Marek. 2023. Feminatywy oraz inne spory o słowa. Próba diagnozy i propozycje pozytywne [Feminatives and other disputed words. Attempt at diagnosis and positive suggestions]. *Socjolingwistyka* 37(1). 345-368.

McConnell, Allen R. & Fazio, Russel H. 1996. Women as men and people: Effects of gender-marked language. *Personality and Social Psychology Bulletin* 22(10). 1004-1013. https://doi.org/10.1177/01461672962210003.

Małocha-Krupa, Agnieszka. 2018. *Feminatywum w uwikłaniach językowo-kulturowych* [The Feminative in linguistic and cultural entanglements]. Wrocław: Oficyna Wydawnicza ATUT.

Małocha-Krupa, Agnieszka. 2021. Feminine personal nouns in the Polish language. Derivational and lexicographical issues. *Lexikos* 31(1). 101-118. https://doi.org/10.5788/31-1-1630

Osińska-Szymańska, Maja & Preuhs, Katarzyna. 2022. Polityczność feminatywów [The political nature of feminatives]. *Studia Polityczne* 50(3). 11-28. https://doi.org/10.35757/STP.2022.50.3.01

Pauwels, Anne. 1998. *Women changing language*. London: Longman.

Pauwels, Anne. 1999. Feminist language planning: Has it been worthwhile? *Linguistik Online* 2(1). https://doi.org/10.13092/lo.2.1043

Rodek, Ewa. 2023. Rzeczowniki żeńskoosobowe zakończone na *-yni/-ini, -ica, -iczka, -aczka, -anka, -arka* w XVII i XVIII wieku (na materiale z Elektronicznego Korpusu Tekstów Polskich z XVII i XVIII w.) [Feminine nouns ending in *-yni /-ini, -ica, -iczka, -aczka, -anka, -arka* in the 17th and 18th centuries (on materials from the Electronic Corpus of Polish Texts from the 17th and 18th centuries)]. *Prace Filologiczne* 78. 339-360.

Satoła-Staśkowiak, Joanna & Sosnowski, Wojciech. 2019. O problematyce kobiecej na płaszczyźnie językowej w kontekście zagadnienia równych szans w Polsce, Rosji i Bułgarii [On women's issues in the linguistic sphere from the perspective of equality of opportunity in Poland, Russia and Bulgaria]. *Socjolingwistyka* 33. 127-147.

Sieczkowski, Andrzej. 1934. Co piszą o języku. *Poradnik Językowy* 3. 45-48.

Skudrzyk, Aldona. 2020. Od prorokini do marszałkini, czyli o derywacji w miarę potrzeb [From prophetess to marshal, i.e. on derivation if needed]. *Annales Universitatis Paedagogicae Cracoviensis. Studia Linguistica* 15. 247-54. https://doi.org/10.24917/20831765.15.20

Smith, Jeremy J. 2023. *Essentials of Early English. An Introduction to Old, Middle, and Early Modern English*. New York: Routledge.

Szpyra-Kozłowska, Jolanta. 2019. Premiera, premierka czy pani premier? Nowe nazwy żeńskie i ograniczenia w ich tworzeniu w świetle badania ankietowego [*Premiera, premierka* or *pani premier*? New female names and restrictions on their creation in light of the survey]. *Język Polski* 99(2). 22-40. https://doi.org/10.31286/JP.99.2.2

Szpyra-Kozłowska, Jolanta. 2021. *Nianiek, ministra i japonki. Eseje o języku i płci*. Kraków: Universitas.

Szpyra-Kozłowska, Jolanta & Laidler, Kateryna. 2022. Czy nazwy żeńskie typu architektka, chirurżka i adiunktka są trudne do wymówienia? Badanie empiryczne [Are feminine nouns such as *architektka* 'female architect,' *chirurżka* 'female surgeon' and *adiunktka* 'female assistant professor' difficult to pronounce? An empirical study]. *Język Polski* 102(4). 37-54.

Wtorkowska, Maria. 2020. Gender, sex, and society. On the formation and usage of feminine forms in Polish and Slovenian. *Acta Universitatis Lodziensis. Kształcenie Polonistyczne Cudzoziemców* 27. 61–80. https://doi.org/10.18778/0860-6587.27.02

Wtorkowska, Maria. 2019. O żeńskich formach nazw zawodów, tytułów i stanowisk w języku polskim [On the feminine forms of the names of professions, titles and positions in Polish]. *Slavistična Revija* 67(2). 223-232. https://srl.si/ojs/srl/article/view/2019-2-1-10

This article was first published at https://www.lacussquare.org

LACUS FORUM

Journal of the Linguistic Association of Canada and the United States

VOLUME 50 NUMBER 3, 2025

CONTENTS

11. A complex approach to rhythm in a minority French community 1
Svetlana Kaminskaïa & Luke Hagar

12. Grammar as a logic of labeling 15
Jacques Lamarche

13. The voice of the unheard: A linguo-cultural world-view of Canada in Polish immigrants' memoirs 28
Maciej Laskowski

14. It's not the glass ceiling—it's the dirt floor. Socio-discursive representations of women from San Juan (Argentina) as workers 40
Tatiana Marisel Pizarro

15. Two birds one stone—idioms across languages and the language brokering effect 51
Joyce Zeng

LACUS Forum **50:3.1-14 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

A complex approach to rhythm in a minority French community

Svetlana Kaminskaïa[1] & Luke Hagar[2]
[1]*University of Waterloo,* [2]*McGill University*

Abstract: Ontario French, spoken in minority settings, often exhibits irregular variation attributable to contact with English. This study focuses on the phonetic rhythm of Windsor French to examine sociolinguistic variation by taking into consideration a series of external and internal factors: age, gender, language use, style, length of utterances, and articulation rate. Using vocalic rhythm metrics – %V, varcoV, and Vnpvi – and linear mixed-effects regression models, we found that style, utterance length, articulation rate, and age consistently affect rhythm. Overall, a more French-like rhythm is associated with reading style and the speech of older participants. Additionally, shorter utterances and faster speech also tend to exhibit such patterns. Overall, rhythmic variation in minority Ontario French seems to show some irregularities and should be further studied in a comparative study with a majority variety.

Keywords: minority French in contact, phonetic rhythm, articulation rate, utterance length, style, social factors

Languages: French, Ontario French in contact with English

1. Introduction

IN CONTACT SETTINGS, more sources of language variation emerge, and it may follow patterns divergent from patterns of variation in non-contact (or majority) situations. In Ontario French, variation is often irregular and unstable – a phenomenon known as sociolinguistic discontinuity (Mougeon & Nadasdi 1998). It affects both linguistic forms and their uses and is often attributable to speakers' linguistic dominance, or language use/restriction (Mougeon & Beniak 1991). Besides the concept of language use, comparisons with majority languages, other contact varieties, and French L2 have also allowed researchers to assess whether variation is irregular and due to the effect of contact with English. The effect of contact has also been observed through speakers' age: younger speakers, being more socially and geographically mobile, engage more intensely with the dominant language, culture, and media. As a result, their speech often reflects a greater influence from the dominant language.

Examples of the effect of English on Ontario French in the speech of restricted or younger speakers have been found in all aspects of language: simplification of verbal paradigms, borrowings, syntactic and semantic transfers, convergences, code-switches, loss of vernacular forms, reduction of assibilation, difficulty in dropping schwa and realizing liaisons, and the use of

falling non-final pitch contours at the expense of the rising ones (see among others: Mougeon & Beniak 1991; Nadasdi 2005; Poiré et al. 2010; Grimm & Nadasdi 2011; Kaminskaïa 2013; Poplack 2018; Blondeau et al. 2019; Mougeon et al. 2022).

As for rhythm, while early studies noted irregularities in Ontario French compared to standard European French (Robinson 1968), recent analyses have found no significant effect of English on its surface pattern – phonetic rhythm. This holds true not only for majority French-speaking communities such as Hearst or Kapuskasing but also for medium-sized communities like North Bay and Timmins, as well as minority varieties (e.g. Windsor) and French L2 learners (Tennant 2011; Kaminskaïa et al. 2016; Tennant & Holmes 2023). These studies show that in different varieties of Ontario French, the measurements that are used to quantify and compare phonetic rhythm (see Section 2) fall within the range of those observed in European and Quebec varieties.

Considering that these measurements represent the *form* that phonetic rhythm manifests and seem unaffected by the contact situation, we ask whether their variation – or the *use* of rhythm – will exhibit irregularities across groups of speakers and different styles. Building on our previous work (Kaminskaïa 2016; 2023), we also propose a more comprehensive analysis that includes language use and internal factors, and aim to determine whether the external factors will yield to internal factors, thus indicating irregular variation (Thomas 1994).

In what follows, we will first provide a brief overview of approaches to classifying phonetic rhythm, methods for its quantification, and studies of its variation. This will lead to our research questions and hypotheses, methodology, and results. A brief discussion will conclude our paper.

2. Phonetic rhythm and its quantification

Phonetic (or prosodic) rhythm, also referred to as *rhythmicity* or *timing*, refers to the rhythm of speech, with its acoustic manifestation being the duration of speech units. The perceived periodicity between these units led to the impressionistic classification of languages into stress-timed (e.g. English), syllable-timed (e.g. French), and mora-timed (e.g. Japanese) (Pike 1945; Abercrombie 1967). Since then, many researchers have attempted to classify languages into these rhythmic categories, with various degrees of success (e.g. Wenk & Wioland 1982 for French). Over time, this traditional classification has been challenged, leading to the concept of a rhythmic continuum. Within this approach, languages have a more or less syllable- or stress-timed rhythmicity, depending on their phonological properties – such as the presence or absence of vowel reduction and the richness or simplicity of syllabic structures (Dauer 1983). These phonological properties affect vowel and consonant durations, and their effects can be cumulative. The measurements of these durations and the comparison of their variability gave rise to the development of rhythm metrics. The most widely used traditional vocalic and consonantal rhythm metrics are listed below (Low 1998; Ramus et al. 1999; Grabe & Low 2002; Ferragne & Pellegrino 2004; Dellwo 2006):

- %V: the proportion of vocalic intervals relative to the consonantal ones
- ΔV and ΔC: standard deviations of the interval durations
- varcoV and varcoC: standard deviations of the vocalic interval durations normalized for articulation rate (100*(ΔV or ΔC)/Ave duration)
- Vrpvi and Crpvi: average of durational differences between adjacent vocalic or consonantal intervals

- Vnpvi and Cnpvi: the rate-normalized equivalent of Vrpvi and Crpvi, calculated as the average from the differences between the adjacent intervals divided by their average duration, and multiplied by 100

A more syllable-timed rhythm is characterized by a higher %V, indicating less consonantal material, along with lower values in other metrics, reflecting the absence of vowel reduction and simpler syllabic structures.

Various rhythm metrics have proven useful in distinguishing regional varieties and identifying the effect of language contact (e.g. Meftah et al. 2013 for Arabic; Clopper & Smiljanic 2015 for American English; Bradshaw et al. 2020 for British English; Michnowicz & Hyler 2020 for Spanish; Morand et al. 2020 for German). However, relatively few studies have examined rhythmic variation across styles, gender, age groups, or social classes. For instance, Arvaniti (2012) and Morand et al. (2022) found that conversational speech tends to be less syllable-timed than text reading; Young (2022) reported that older speakers and those from upper classes exhibit less syllable-timed rhythms compared to younger speakers and those from lower social classes. Similarly, Meftah et al. (2013) observed that women tend to have a more syllable-timed rhythm than men.

There are even fewer studies that have considered linguistic factors as predictors of phonetic rhythm. One example is Clopper & Smiljanic's (2015) analysis, where articulation rate was included in the models assessing variation of an array of rhythm metrics relative to various factors. Another example is Young's (2022) study, where a series of linguistic factors, including articulation rate, were taken into consideration for the analysis of Vnpvi. These are also two studies that examined rhythm metrics not as individual averages (one per speaker) but as a series of values at the level of observation (i.e. values observed at the level of each utterance and of each vowel pair respectively). While Young (2022) found that rate had a significant effect on Vnpvi (p. 1056), Clopper & Smiljanic (2015) concluded that Vnpvi and varcoV appeared "effectively normalized for articulation rate" (p. 10).

2.1 Variation of phonetic rhythm in French

In French, regional rhythmic variation has been explored in many analyses. Some studies on European and African French have identified regional and contact effects, while others have not. For example, Avanzi et al. (2012) and Obin et al. (2012) found that in text readings, "peripheral" varieties and contact varieties spoken in Switzerland, Central Africa, and Senegal exhibit less syllable-timed rhythm compared to more central or "standard" varieties. However, Meisenburg (2013) found no evidence of contact effect in her comparison of text readings by Occitan speakers with those by speakers of Southern and Northern French. Instead, the comparison of Occitan spontaneous speech with text readings revealed stylistic differences: conversations tended to be less syllable-timed than readings. Notably, gender and age did not seem to influence rhythmic patterns in European French, as previous studies reported no significant differences despite including male and female participants and speakers of varying ages.

In Canadian Acadian varieties of French, Cichocki et al. (2013; 2020) found that region, style, age, and gender all influenced rhythmic variation. However, studies on Canadian Laurentian varieties, that include Ontario dialects, did not find significant regional or contact effects (Kaminskaïa et al. 2016; Tennant & Holmes 2023). Instead, gender, age, and style (Kaminskaïa 2016) played a role: women, older participants, and text readings showed a more syllable-timed rhythm.

These analyses did not take into consideration internal factors, and, as most of the studies before, relied on individual averages (i.e. one metric value per speaker) rather than analyzing a series of metrics across all utterances produced by each speaker. In our further analysis (Kaminskaïa 2023), we examined two metrics – varcoV and Vnpvi – at the level of utterance and included articulation rate and utterance length (in number of syllables). We found that gender was not significant anymore and that internal factors were each significantly associated with one of the metrics: articulation rate had a negative association with Vnpvi, and utterance length had a positive association with varcoV.

2.2 Objectives

Previous analyses of French prosodic rhythm provided valuable insights into rhythmic variation in a contact setting. However, they did not take into account language use as an explicit variable. Here, we build on this and on our own previous work and propose an analysis that examines rhythm, as expressed by three vocalic metrics – %V, varcoV, and Vnpvi. Vocalic metrics in general have been reported to best capture variation in a contact setting in French (Guilbault 2002) and in English (Bradshaw et al. 2020). These three metrics, in particular, have been reported to have the most discriminative power and to be the most robust (Wiget et al. 2010; Prieto et al. 2012, among others). We analyze rhythm at the utterance level and consider multiple factors, including style, age, gender, language use, articulation rate, and utterance length, to offer a more comprehensive understanding of rhythmic variation in Ontario French.

2.3 Research questions and hypotheses

The primary research question concerns the extent of rhythmic variation and whether this new analysis will confirm our previous findings regarding the importance of external factors or if linguistic factors will dominate, indicating irregular variation (see Thomas 1994). A further question is whether both age and language use will prove significant, given that previous studies have normally highlighted age or density of the francophone population in the examined community, to demonstrate the effect of contact (Kaminskaïa 2014; Tennant & Holmes 2023).

We hypothesize that both internal and external factors will be important. Regarding external factors, we expect our previous findings to be confirmed: older participants, women, and text readings will show a more regular (syllable-timed) rhythm. We also anticipate an interaction between age and style, with younger participants producing less regular rhythm in spontaneous speech. Additionally, language use is expected to align with age in our data. As for internal factors, consistent with earlier research, as articulation rate increases, we expect to observe a trend towards a more syllable-timed pattern. Conversely, as the number of syllables in the utterance grows, we anticipate a less syllable-timed rhythm.

3. Methodology

3.1 Data and participants

The data for this study come from the *Phonologie du français contemporain (PFC)* project (Durand et al. 2009), specifically the Windsor corpus (Poiré & Kelly 2003). We use text readings and spontaneous conversations (dialogues with family members or friends). All participants (Table 1) are native French speakers, with ages ranging from 17 to 84; there are 6 males and 6

females.

Language use codes were assigned based on participants' self-reported French language practices: « 0 » corresponds to the least restricted speakers who predominantly use French in daily interactions, « 1 » to semi-restricted speakers, and « 2 » to the most restricted speakers (Poiré 2009). Each language-use group contains four speakers. However, these codes were assigned after data collection, resulting in some imbalance between language restriction and gender, or age. Still, the dataset remains sufficiently balanced to include all relevant factors in the analysis.

Social characteristics such as socioeconomic background were not considered, as social distinctions in this minority setting are minimal and are not expected to significantly influence linguistic behaviour (Thomas 1994).

Table 1 : Participants (adapted from Poiré 2009: 159)

Code	Gender	Age	Language Use	Occupation	Education	Relationships
Coadp1	F	43	1	Housewife (teacher)	B.A.	Spouses
Coacd1	M	41	1	Technician	College	
Coahc1	F	65	0	Retired (farmer, teacher)	Teacher's college	Spouses and their son
Coarc1	M	66	0	Retired (farmer)	Elementary	
Coamc1	M	33	2	Electrician	College	
Coars1	M	56	2	Industry	High school	Father and son
Coams1	M	21	1	Undergradute student	University student	
Coals1	F	74	2	Retired (bank employee)	High school	Spouses
Coars2	M	74	2	Retired (railway worker)	High school	
Coams2	F	84	0	Retired (bank manager)	College	Grandmother and granddaughter, friend
Coarb1	F	17	1	Student	High school	
Coabb1	F	42	0	Nurse	B.A.	

3.2 Analyses

Acoustic analysis of the recordings was conducted using *Praat* (Boersma & Weenink 2022). Initial segmentation was performed with *EasyAlign* (Goldman 2011) and followed by manual verification. To ensure balance between the two styles, we analyzed a similar number of utterances: 789 from text readings and 872 from conversations. We excluded all dysfluencies and utterances containing fewer than two vocalic intervals necessary for calculating the metrics. Utterances were defined as stretches of speech occurring between two silent pauses of 80 milliseconds or more, referred to as Inter-Pause Intervals (IPIs). Using *Correlatore* (Mairano 2011), we obtained values of series of metrics for each IPI; however, we focus only on the three most stable and reliable metrics: %V, Vnpvi, and varcoV. We also recorded the length of each IPI in terms of the number of syllables (NmbSyll) and calculated its articulation rate (ArtRate) in syllables per second.

To examine the effects of internal and external factors on each metric, we applied linear mixed-effects models using the *nlme* package in *R* (Pinheiro et al. 2017; R Core Team 2020). The

values of the metrics were the dependent variables; ArtRate and NmbSyll were included as covariates; and Style (binary variable), Gender (binary variable), Age (treated as a continuous variable), and Language use (a three-level variable) were independent factors. Models also included a random intercept for each speaker.

We applied the necessary data transformations to the values of the metrics, ArtRate, and NmbSyll to meet the assumptions required for the models. NmbSyll, ArtRate, and Age were centered at their median values (4, 4.64, and 43, respectively).

It is noteworthy that including ArtRate in the analysis of rhythm metrics – especially %V – is unconventional. This decision was informed by observations regarding the sensitivity of metrics to articulation rate in general (Arvaniti 2012), as well as the observation that vowel durations are compressed more than consonant durations at faster speech rates in French (Fletcher 1987: 964). Furthermore, incorporating ArtRate improved the stability of the model for %V, which further validates our methodological approach.

4. Results

We will begin by making overall observations of potentially significant trends for each metric. This will involve plotting each metric, on the Y-axis, against continuous variables – such as Age, ArtRate, or NmbSyll – on the X-axis, while splitting the data by Gender, Style, or Language use. Following this, we will assess whether the observed trends are supported by the best-fitting linear mixed-effects model.

4.1 %V

The values for the %V metric showed moderate positive skewness and were therefore transformed using the square root function. Similarly, ArtRate and NmbSyll underwent this transformation to maintain a linear relationship between the continuous variables and the response.

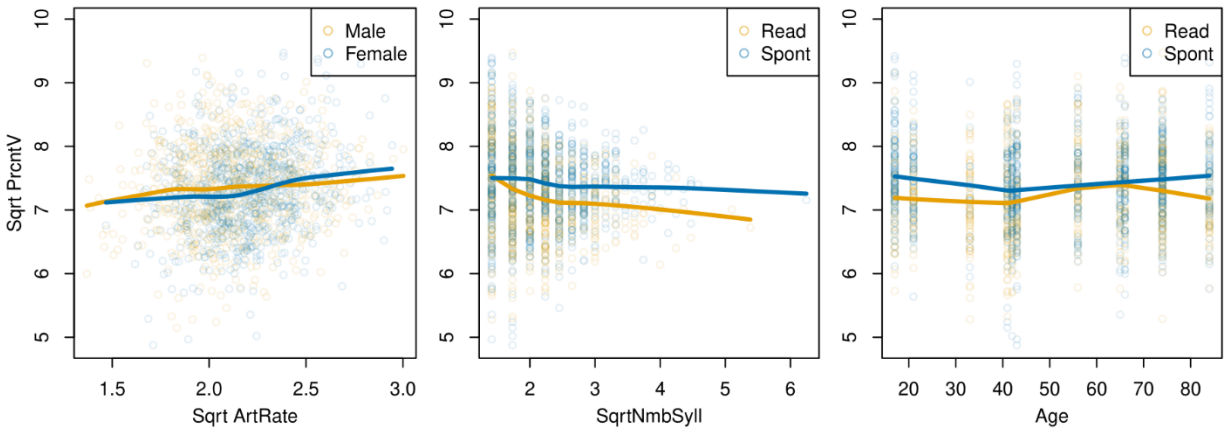


Figure 1: Potentially important relationships between sqrt(%V) and the assessed factors

As illustrated in Figure 1, %V values increase when ArtRate increases (left panel), and they decrease when the number of syllables in an IPI grows (middle panel), particularly in the reading style. The split between reading and spontaneous speech suggests that this is an important factor as spontaneous speech tends to have greater %V values. At the same time, %V remains stable

across different ages and among both males and female participants (right panel).

The best-fitting mixed-effects model (Table 2) confirms that ArtRate ($p < .001$) and the interaction between Style and NmbSyll ($p < .001$) are significantly associated with %V. So, rhythm becomes more syllable-timed as ArtRate increases, and it becomes less syllable-timed as utterances' NmbSyll grows, especially in the reading style. Additionally, both Style and NmbSyll are individually significant ($p \leq .047$). Gender, Language use, Age, and its interaction with Style were not significant ($p \geq .41$).

Table 2: Estimated parameters (with standard error) of linear mixed model of sqrt(%V)

Effect	Estimate	Std Error	*p*-value
Fixed effects			
Intercept	7.314	0.044	<.001
Style: Spont vs. Read	0.073	0.037	.047
sqrt(ArtRate)	0.452	0.074	<.001
sqrt(NmbSyll)	-0.270	0.040	<.001
sqrt(NmbSyll)*Style	0.175	0.051	<.001
Random effects			
Intercept	–	0.124	–

4.2 Vnpvi

The values of the Vnpvi metric had moderate right skewness, prompting us to apply a square root transformation to achieve a more symmetrical data distribution. To meet model assumptions, we similarly transformed ArtRate and NmbSyll values using the square root function.

As illustrated in Figure 2, Vnpvi appears to decrease overall with both Age and ArtRate. In the left panel, younger participants show a distinction between Vnpvi values for text readings (lower values) and spontaneous speech (higher values). The middle panel suggests a potential difference between male and female participants, while the right panel indicates a possible effect of Language use at slower articulation rates.

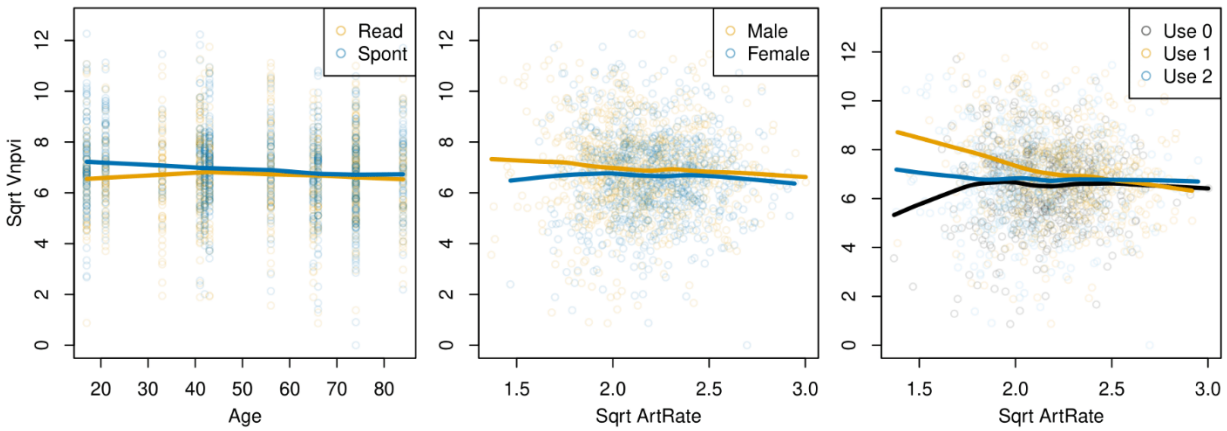


Figure 2: Potentially important relationships between sqrt(Vnpvi) and the assessed factors

However, mixed-effects modeling did not show any significant effects for Gender, Language use, or interactions between factors ($p \geq .06$). In contrast, the best model (Table 3) revealed the

effects of Style, ArtRate, and Age ($p \leq .03$): in our data, rhythm becomes more syllable-timed as both age and articulation rate increase, and spontaneous speech appears less syllable-timed.

Table 3: Estimated parameters (with standard error) of linear mixed model of sqrt(Vnpvi)

Effect	Estimate	Std Error	*p*-value
Fixed effects			
Intercept	6.685	0.091	<.001
Style: Spont vs. Read	0.339	0.092	<.001
sqrt(ArtRate)	-0.609	0.186	.001
Age	-0.009	0.003	.03
Random effects			
Intercept	–	0.203	–

4.3 varcoV

No transformations were applied to the varcoV metric; however, square root transformations were utilized for the NmbSyll and ArtRate due to their skewness.

The left panel in Figure 3 indicates that while varcoV generally decreases with age, values for older participants overlap across both styles, whereas in younger participants, spontaneous speech has higher varcoV values. This suggests a potential interaction between Age and Style. Although ArtRate does not appear to influence varcoV overall, there may be differences between male and female speakers (middle panel), and Language use could also be important (right panel): male participants and semi-restricted speakers (Use 1) tend to have higher varcoV values. At the same time, the increase in the number of syllables within the IPI contributes to the increase of the varcoV values (right panel).

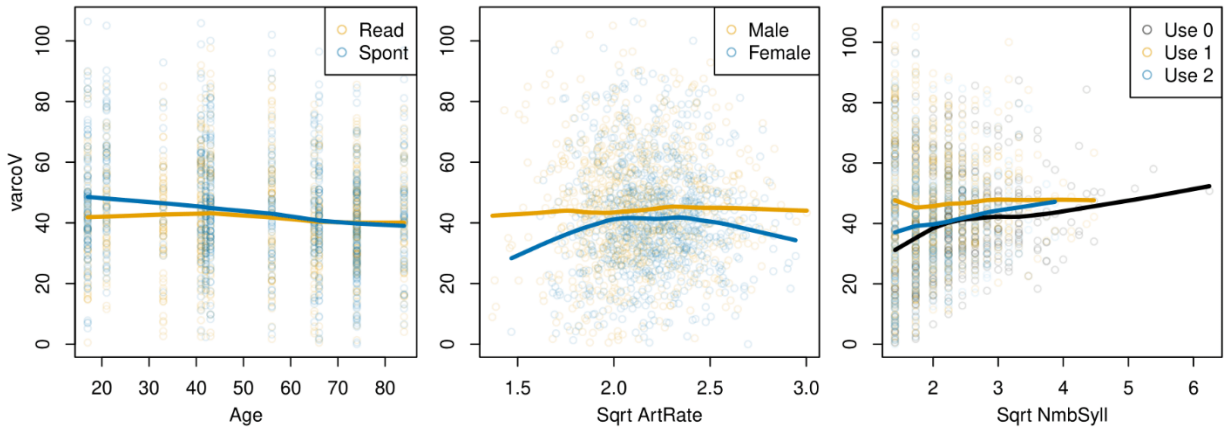


Figure 3: Potentially important relationships between varcoV and the assessed factors

Significance of NmbSyll and that of the interaction between Age and Style were confirmed ($p \leq .008$) (Table 4). However, the effects of Gender and Language use were not significant ($p \geq .12$), and neither was ArtRate ($p = .21$). In other words, the spontaneous speech of younger speakers appears less syllable-timed, and the vocalic interval durations show more variability in longer utterances.

Table 4: Estimated parameters (with standard error) of linear mixed model of varcoV

Effect	Estimate	Std Error	*p*-value
Fixed effects			
Intercept	41.937	1.237	<.001
Style: Spont vs. Read	3.334	1.000	<.001
Age	-0.053	0.055	.361
Age*Style	-0.117	0.044	.008
sqrt(NmbSyll)	2.643	0.717	<.001
Random effects			
Intercept	–	3.231	–

5. Discussion and conclusions

In this paper, we examined rhythmic variation in a minority French community by adopting a comprehensive approach that considers both external and internal factors. The primary objective was to assess the extent of rhythmic variation and determine whether it exhibits irregularity, which, according to Thomas (1994), would be indicated by a predominance of linguistic factors. The secondary goal was to explore how Age and Language use contribute to the characterization of rhythmic variation in our data. We hypothesized that the latter factors would have a parallel effect on rhythm and expected to observe a more syllable-timed pattern in text readings, in females' speech, in faster speech, and in shorter utterances.

Our results partially confirm our hypotheses regarding the extent of variation. Compared to Kaminskaïa (2016), fewer external effects were found to be significant in this study, and those that were significant showed interactions. In respect to the first objective, our analysis indicated that two internal factors (ArtRate and NmbSyll) and two external factors (Style and Age) consistently influence phonetic rhythm in our data. More specifically, the increase of ArtRate makes rhythm more syllable-timed, while lengthening of IPIs makes it less syllable-timed, a trend that is more pronounced in text readings. At the same time, spontaneous speech is less syllable-timed, especially in younger speakers. Gender did not have a significant effect on rhythm in this analysis.

The findings related to Style are consistent with studies of other languages and varieties of French that were presented above. This suggests that speakers pay similar attention to style differentiation, and in many languages reading style exhibits a more regular rhythm. Regarding Age, current results diverge from our previous findings (Kaminskaïa 2016), where the effect of age was observed in both styles. This discrepancy is likely attributable to methodological differences: in the present study, age was treated as a continuous variable rather than a binary category, and the analysis was conducted at the level of the IPI. This approach contributed to the interaction between Style and Age, wherein younger participants demonstrated a less syllable-timed rhythm in spontaneous speech compared to their own text readings, while older speakers maintained more consistent patterns across both styles. This finding may indicate an ongoing change among younger participants or suggest that older speakers have difficulty differentiating between styles. Further analysis of additional prosodic aspects could provide more insights into this observation. We note, however, that stylistic variation is not the same across the age continuum. Gender was not found to be significant in the present analysis, which aligns with sociolinguistic practice when more (potentially more influential) factors are considered and with Thomas' (1994) observation about external factors yielding before the internal ones in a minority setting. These results also align with Kaminskaïa (2023), where Language use was not taken into

consideration.

Regarding the effect of internal factors, both variables tested were found to be significant. ArtRate demonstrated significance for Vnpvi and %V but not for varcoV, reinforcing the latter's status as the most stable metric. For Vnpvi, our results are consistent with Young (2022) but do not align with Clopper & Smiljanic's (2015) study. Young calculated rate as the average of consecutive vowel durations, which complicates the comparison with our study. As for the divergence of our results from Clopper & Smiljanic's, a considerably slower rate of Windsor speakers (mean of 4.64 syll/s) compared to the American English participants (regional means from 5.36 syll/s to 5.72 syll/s) can be behind the significant association between ArtRate and Vnpvi in our analysis (also, see below). At slower rates, durational variability must not be sufficiently normalized for Vnpvi. Significant association between ArtRate and %V validated the inclusion of this internal factor in the analysis of this particular metric. The direction of the effect (%V decreases as ArtRate increases) suggests that anticipatory shortening (Bishop & Kim 2018) affects vocalic and intervocalic intervals differently, thus supporting that in French, at faster articulation rates, vocalic durations get compressed more than consonantal durations (Fletcher 1987).

The impact of NmbSyll was observed on varcoV and, in combination with Style, on %V. We believe that for %V this relationship may be linked to syllabic complexity: longer IPIs have a richer array of syllable structures, which affects the ratio between vocalic and intervocalic intervals. Interaction with Style suggests that in text readings syllabic typology may be more complex, not only due to sentence composition but also due to a more careful articulation and non-application of such processes as consonant cluster simplification in a more formal style, or resyllabification.

%V was the only metric affected by both linguistic variables ArtRate and NmbSyll. As discussed above, syllabic typology and anticipatory shortening may underlie these effects. Vnpvi had a significant association with ArtRate but not with NmbSyll, while varcoV had one with NmbSyll but not with ArtRate. As observed by Wiget et al. (2010: 1565), Vnpvi is sensitive to the alternation of stressed and unstressed syllables. This translates into the frequency of occurrence of stress, i.e. the size of the stress group or Accentual Phrase (AP). The overall average length of APs in our data is 3.65 syllables. The most important acoustic correlate of stress in French is duration – the obligatory stress is AP-final, and it is marked by durational lengthening. So, both size of the AP and the nature of stress create a context where the articulation rate will affect durational variability of consecutive intervals (hence, the effect of ArtRate on Vnpvi). For varcoV, alternation between stressed/unstressed elements is not relevant; instead, durations of all intervals matter (ibid.). The more intervals there are, the greater their expected durational variability. This explains the significant association of varcoV with NmbSyll rather than with ArtRate.

Finally, addressing the third research question concerning the importance of both Age and Language use, age provided a better account of rhythm variation in our dataset. This is obviously due to the nature of our data. A larger and more balanced corpus could help us better understand the relationship between these factors.

All in all, a more comprehensive approach and a more complex analysis nuanced our previous findings: after taking into consideration internal factors and examining rhythm at the level of IPI, the analysis revealed the effect of both internal factors and an interaction between Age and Style, but removed the effect of Gender. According to Thomas (1994), this suggests that variation in our data is not quite regular. However, the predominance of linguistic factors over the external ones has also been observed in languages in a non-contact situation (Tanner et al. 2024). This calls for a comparison of Windsor French with majority varieties, such as Quebec French, to understand if

the observed variation is an indication of an on-going change in Ontario French or a more general characteristic of Laurentian French.

Acknowledgements

We extend our gratitude to the participants who generously contributed to the PFC project, Windsor corpus. We also thank François Poiré for granting us access to the data. We thank two anonymous reviewers for their comments on an earlier version of this paper. This study was funded by the Social Sciences and Humanities Research Council of Canada (Insight Grant #1675-50656-10104 to S. Kaminskaïa).

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40. 351-373.

Avanzi, Mathieu & Obin, Nicolas & Bardiaux, Alice & Bordal, Guri. 2012. Speech prosody of French regional varieties. In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel (eds.), *Speech Prosody* 6, 603-606. Shanghai, Chine.

Bishop, Jason & Kim, Boram. 2018. Anticipatory shortening: Articulation rate, phrase length, and lookahead in speech production. *Speech Prosody 2018*, 235-239. In Klessa, Katarzyna & Bachan, Jolanta & Wagner, Agnieszka & Karpiński, Maciej & Śledziński, Daniel (eds.) *Proceedings of the 9th International Conference on Speech Prosody 2018*, 235-239. Poznan, Poland.

Blondeau, Hélène & Mougeon, Raymond & Tremblay, Mireille. 2019. Analyse comparative de *ça fait que, alors, donc* et *so* à Montréal et à Welland: Mutations sociales, convergences, divergences en français laurentien. *Journal of French Language Studies* 29(1). 35-65.

Boersma, Paul & Weenink, David. 2022. *Praat: doing Phonetics by Computer.* (https://www.fon.hum.uva.nl/praat/)

Bradshaw, Leah & Hughes, Vincent & Chodroff, Eleanor. 2020. Investigating the forensic applications of global and local temporal representations of speech for dialect discrimination. *Speech Prosody 2020,* 635-639. ISCA.

Cichocki, Wladyslaw & Selouani, Sid-Ahmed & Ayed, Alaidine B. & Boudreau, Catherine & Perreault, Yves. 2013. Variation of rhythm metrics in regional varieties of Acadian French. Montreal, Canada. *Proceedings of Meetings on Acoustics,* 19: 035082. ICA 2013 Montréal, June 2013.

Cichocki, Wladyslaw & Selouani, Sid-Ahmed & Perreault, Yves. 2020. Prosodic rhythm in regional varieties of French in New Brunswick (Canada). In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences,* 2310-2314. Australasian Speech Science and Technology Association Inc., and International Phonetic Association.

Clopper, Cynthia G. & Smiljanic, Rajka. 2015. Regional variation in temporal organization in American English. *Journal of Phonetics* 49. 1-15.

Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1). 51-62.

Dellwo, Volker. 2006. Rhythm and speech rate: A variation coefficient for deltaC. In Karnowski, Pawel & Szigeti, Imre (eds.), *Sprache und Sprachverarbeitung: Akten des 38. Linguistischen Kolloquiums in Piliscsaba 2003* [Language and language-processing: Proceedings of the 38th Linguistics Colloquium, Piliscsaba 2003], *Linguistik International* 15, 231-241. Frankfurt am Main, Germany: Peter Lang Publishing Group.

Durand, Jacques & Laks, Bernard & Lyche, Chantal. 2009. Le projet PFC : Une source de données primaires structurées. *Phonologie, variation et accents du français*, 19-61. Hermès.

Ferragne, Emmanuel & Pellegrino, François. 2004. A comparative account of the suprasegmental and rhythmic features of British English dialects. *Proceedings of "Modelisations pour l'Identification des Langues,"* 121-126. Paris, France.

Fletcher, Janet. 1987. Some micro and macro effects of tempo change on timing in French. *Linguistics* 25(5). 951-968.

Goldman, Jean-Philippe. 2011. EasyAlign: A friendly automatic phonetic alignment tool under Praat. *Proceedings of the 12th Annual Conference of the International Speech Communication Association 2011*, 3233-3236.

Grabe, Esther & Low, Ee Ling. 2002. Durational variability in speech and the Rhythm Class Hypothesis. In Gussenhoven, Carlos & Warner, Natasha (eds.), *Laboratory Phonology* 7, 515-546.

Grimm, Rick & Nadasdi, Terry. 2011. The future of Ontario French. *Journal of French Language Studies* 21(2). 173-189.

Guilbault, Christian. 2002. *The acquisition of French rhythm by English second language learners*. Edmonton: University of Alberta.

Kaminskaïa, Svetlana. 2013. Intonation of Ontario French in a minority setting: A study comparing two age groups. *The French Review* 87(2). 119-139.

Kaminskaïa, Svetlana. 2014. Quantifying rhythm in minority French. *The Italian Journal of Linguistics* [Rivista di Linguistica] 26(1). 93-133.

Kaminskaïa, Svetlana. 2016. Interplay of sociolinguistic factors in rhythmic variation in a minority French dialect. *Proceedings of the 8th International Conference on Speech Prosody 2016*, 1191-1195.

Kaminskaïa, Svetlana. 2023. Sur la variabilité du rythme phonétique en français ontarien minoritaire. In Barysevich, Alena & Arroyas, Frédérique & Irvine, Margot (eds.), *Perspectives sociolinguistiques variationnistes du français en situation de contact des langues. Mélanges en hommage à Alain Thomas. Nouvelle Revue Synergies Canada* 17. https://doi.org/10.21083/nrsc.v2023i17.7301

Kaminskaïa, Svetlana & Tennant, Jeff & Russell, Alexander. 2016. Prosodic rhythm in Ontario French. *Journal of French Language Studies* 26(2). 183-208.

Low, Ee Ling. 1998. *Prosodic prominence in Singapore English.* Cambridge: University of Cambridge. (Doctoral dissertation.)

Mairano, Paolo. 2011. *Rhythm typology: Acoustic and perceptive studies*. Turin: Università degli studi di Torino. (Doctoral dissertation.)

Meftah, Ali H. & Selouani, Sid-Ahmed & Alotaibi, Yousef A. 2013. Investigating speaker gender using rhythm metrics in Arabic dialects. *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, 347-350.

Meisenburg, Trudel. 2013. Southern vibes? On rhythmic features of (Midi) French. *Language Sciences* 39. 167-177.

Michnowicz, Jim & Hyler, Alex. 2020. Chapter 2. The changing rhythm of Yucatan Spanish. In Rao, Rajiv (ed.), *Issues in Hispanic and Lusophone Linguistics* 28. 33-62. Amsterdam: John Benjamins Publishing Company.

Morand, Marie-Anne & Bruno, Melissa & Julmi, Nora & Schwab, Sandra & Schmid, Stephan. 2020. Speech rhythm in multiethnolectal Zurich German. *Speech Prosody 2020*, 566-570. ISCA.

Morand, Marie-Anne & Bruno, Melissa & Schwab, Sandra & Schmid, Stephan. 2022. Syllable rate and speech rhythm in multiethnolectal Zurich German: A comparison of speaking styles. *Speech Prosody 2022*, 337-341. ISCA.

Mougeon, Raymond & Beniak, Édouard. 1991. *Linguistic consequences of language contact and restriction: The case of French in Ontario*. Oxford: Oxford University Press.

Mougeon, Raymond & Nadasdi, Terry. 1998. Sociolinguistic discontinuity in minority language communities. *Language* 74(1). 40-55.

Mougeon, Raymond & Rehner, Katherine & Mougeon, Françoise. 2022. Participation in (non)salient linguistic change over the lifespan: An examination of panel speakers' life stories. *Journal of Sociolinguistics* 26(2). 265-286.

Nadasdi, Terry. 2005. Le français en Ontario. In Valdman, Albert & Auger, Julie & Piston-Hatlen, Deborah (eds.), *Le français en Amérique du Nord,* 99-115. Québec: Presses de l'Université Laval.

Obin, Nicolas & Avanzi, Mathieu & Bordal, Guri & Bardiaux, Alice. 2012. Regional variations of speech rhythm in French: "In Search of Lost Times." In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel (eds.), *Speech Prosody 2012*, 406-409. Shanghai, China.

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pinheiro, José & Bates, Douglas & DebRoy, Saikar & Heisterkamp, Siem & van Willigen, Bert & Ranke, Johannes. 2017. *Package 'nlme.' (R.)*

Poiré, François. 2009. Le français canadien en milieu minoritaire: Le cas du Sud-ouest ontarien. In Durand, Jacques & Laks, Bernard & Lyche, Chantal (eds.), *Phonologie, variation et accents du français*. 153-174. Paris: Hermès.

Poiré, François & Kaminskaïa, Svetlana & Tremblay, Remi. 2010. Conséquences du contact avec l'anglais sur la réalisation de la liaison et du schwa en français de Windsor, Canada. *XXVe CILPR Congrès International de Linguistique et de Philologie Romanes: Innsbruck, 3-8 septembre 2007*, 365-374. Innsbruck : De Gruyter.

Poiré, François & Kelly, Stephanie. 2003. Présentation de l'étude du français, langue minoritaire, dans le sud-ouest ontarien dans le cadre du PFC. Maison des sciences de l'homme, Paris, France. (Conference presentation at the *Colloque Phonologie et phonétique du français: données et théorie*, Maison des sciences de l'homme, Paris, France.)

Poplack, Shana. 2018. *Borrowing: Loanwords in the speech community and in the grammar*. New York: Oxford University Press.

Prieto, Pilar & del Mar Vanrell, Maria & Astruc, Lluïsa & Payne, Elinor & Post, Brechtje. 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication* 54. 681-702.

R Core Team. 2020. *R: A language and environment for statistical computing*. (Vienna, Austria.) (https://www.R-project.org/)

Ramus, Franck & Nespor, Marina & Mehler, Jacques. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265-292.

Robinson, Lynda. 1968. Étude du rythme syllabique en français canadien et en français standard. In Léon, Pierre (ed.), *Recherches sur la structure phonique du français canadien* (Studia Phonetica 1), 161-174. Montréal: Didier.

Tanner, James & Sonderegger, Morgan & Stuart-Smith, Jane & Kendall, Tyler & Mielke, Jeff & Dodsworth, Robin & Thomas, Erik. 2024. Exploring the anatomy of articulation rate in spontaneous English speech: Relationships between utterance length effects and social factors. *Proceedings of Interspeech 2024*, 467-471. ISCA.

Tennant, Jeff. 2011. Rythme prosodique et contact des langues dans le français ontarien. In Martineau, France & Nadasdi, Terry (eds.), *Le français en contact. Hommages à Raymond Mougeon*, 355-373. Québec: Presses de l'Université Laval.

Tennant, Jeff & Holmes, Damaris. 2023. Rythme prosodique dans le français et l'anglais des francophones du Nord-Est de l'Ontario. In Barysevich, Alena & Arroyas, Frédérique & Irvine, Margot (eds.), *Perspectives sociolinguistiques variationnistes du français en situation de contact des langues. Mélanges en hommage à Alain Thomas. Nouvelle Revue Synergies Canada* 17. https://doi.org/10.21083/nrsc.v2023i17.7279

Thomas, Alain. 1994. La prononciation du français dans le Moyen-Nord ontarien. *Francophonies d'Amérique* 4. 5-11.

Wenk, Brian J. & Wioland, François. 1982. Is French really syllable timed? *Journal of Phonetics* 10. 193-216.

Wiget, Lukas & White, Laurence & Schuppler, Barbara & Grenon, Izabelle & Rauch, Olesya & Mattys, Sven L. 2010. How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America* 127(3). 1559-1569.

Young, Nathan J. 2022. The sociolectal and stylistic variability of rhythm in Stockholm. *Language and Speech* 65(4). 1034-1070.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:3.15-27 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Grammar as a logic of labeling

Jacques Lamarche
Western University

Abstract: This paper presents an analysis of verb-noun conversion in English in a model of grammar where the lexical primitives of syntax are the phonological matrices of words, essentially formal *identifiers* or *labels*, that are recognized as units of a language because they conventionally identify denotational content assumed to be outside of grammar. In this Logic of Labeling, grammatical values reflect the position of lexical forms in an endocentric audible constituent: a verbal value is defined as the head, and the nominal value as the non-head of such a constituent. A lexeme that is a non-head (a noun) can identify its denotation with a simple form (a term), but as head (a verb), the denotation must be identified by a full audible constituent. Conversion occurs when the formal part of a given lexeme can appear as head or non-head of audible constituents, independently of its lexical denotation. This provides a direct account of transparent conversion cases, where no change in denotation occurs between nominal and verbal uses. Differences in interpretation are then all tied to the position of the lexeme's matrix in audible constituents.

Keywords: conversion, syntax, semantics, linguistic form, constituent

Language: English

1. Introduction

THIS PAPER PRESENTS AN ACCOUNT for verb-noun conversion in English – when a lexeme can be used either as a verb or a noun without a derivational affix – that does not rely on specific grammatical abstractions (zero-morphemes or a priori UG category features) or non-grammatical (external) notions such as discourse functions, denotational content, semantic/thematic roles, and so on. In the proposed approach, the constituent structure where distributional values are defined is made up of the phonological matrices of lexical words: grammatical values then strictly depend on the position of a word's formal part in this structure. This account provides a natural framework for conversion in oral language given that, in principle, any word can appear as a verb or a noun depending on where its matrix appears in the audible constituent structure.[1]

In this model of grammar, called Logic of Labeling, the lexical phonological matrices of words are uncategorized identifying labels: they are sequences of phonological instructions recognized

[1] Evidently, the notion of audible constituent can only pertain to oral language. Although I cannot in the current paper deal with sign language, the expectation under the current approach is that the multiple modalities of sign language also obey the endocentricity (Section 2.3) and alignment (Section 3.1) principles central to the current account.

as units of speakers' language because these formal objects identify (i.e. apply to) arbitrary denotations that are assumed to be outside of grammar. Grammar – understood as a subcomponent of language, not equivalent to language – is a labeling device that generates phonemes and combines them to create simple labels. Simple labels are combined into complex labels, that I call *audible constituents*. In language use, these simple and complex labels are applied to realities that are outside of grammar. This model thus presupposes i) a pre-existing background composed of a world (i.e. Semantics) shared by speakers ii) knowledge of this world, which must be public, equally available to all in the sense of Fodor (1998), so that lexical conventions can be established and learned. Significant aspects of a specific language like English are thus outside of the labeling function of grammar, as this function is to provide the formal material used to identify pre-existing external realities – the lexical denotations of words.

The hypothesis is that the difference between verbs and nouns lies in the degree of complexity of the audible form that identifies denotations. Using endocentric principles, verbal and nominal values are distinguished, respectively, as the head and non-head positions in constituents. With a nominal non-head, a simple form (or term) is sufficient to identify a denotation: the formal object essentially names the denotation. As head of the audible constituent, however, verbs identify their arbitrary denotation as a full constituent, explicitly stretching this denotation over the whole formal object. This denotation is then understood not just as a name, but as naming an actual situation. Since the account of the relevant values is based on the complexity of the formal object that applies to an arbitrary denotation, rather than being dependent on the denotation itself, the analysis predicts that in principle, a lexical form could appear either as a verb or as a noun regardless of its denotation, which could remain the same across the two uses, that is, a completely transparent conversion. Many cases of conversions in English are transparent in this sense. Whether a lexeme is understood as the verbal head of a proposition (an expression that can have a truth-value) or as one of the participants of a proposition (a noun), is thus strictly a matter of formal (grammatical) analysis, not denotation (i.e. Semantics).

In Section 2, I present the conversion facts and the analysis of the verb-noun dichotomy in terms of endocentric principles applied to audible constituents. The account is formalized in Section 3 using input forms to represent the relevant distinctions. The analysis is summarized in Section 4.

2. V-N conversion in English

Conversion is the process by which a lexeme's category is changed without any derivational marking. The scope of the phenomena is such that, as Lieber (2005: 418) puts it "one might view conversion as a sort of battleground over which various theoretical camps have fought over the years."[2] My objective here is to examine a sub-case of verb-noun conversion – transparent cases, that is, a constant denotation in verbal and nominal uses – and show that all relevant distinctions between the two uses depend on the position of the words in audible constituents, not on lexical or denotation content.

The analysis is informally presented in Section 2.1, using the conversion facts seen with lexemes like *rain, snow, hail*, etc., called precipitation lexemes in Levin & Krejci (2019). Section 2.2 argues that the difference in complexity can be tied to a basic difference in humans'

[2] See Marchand (1960), Chapter V, for a comprehensive overview for English. The issues raised by conversion include: the category of words that undergo conversion and the direction of conversion; the productivity of conversion within and across languages; the semantic differences observed in specific cases; diachronic dimensions of conversion; etc.

understanding of the world. Section 2.3 uses endocentric properties to define the verb-noun dichotomy and Section 2.4 extends the analysis of conversion to other intransitive verbs where conversion is transparent (no meaning changes between verbal and nominal uses).

2.1 Precipitation lexemes in English

Among weather expressions, precipitation lexemes like *rain* have two characteristics relevant for the current discussion. The first, illustrated in (1a), is that their subject is an expletive *it*.[3]

(1) a. *It rains/is raining.*
b. *The rain is cold.*

The second characteristic is that it is perfectly possible for the denotation to be the same in the verbal and nominal uses.[4] For instance, when *rain* is used as a noun with a determiner in (1b), its denotation appears the same as in its verbal version in (1a) (something like ⟦water falling from the sky⟧[5]). For instance, if I utter *it is raining* (verbal *rain*) to describe the weather and then continue with *the rain is really cold* (nominal *rain*), the implication is that the form *rain* in the two utterances denotes the exact same reality in the context.[6] In contrast, if *Mary is hammering a nail* or *is buttering a piece of toast*, or *is canning the jam*, the denominal verbal events necessarily imply more than a hammer, some butter or a can.

This second characteristic is significant in that it suggests that conversion cannot be tied to any difference in the denotation of *rain* because its contribution ⟦water falling from the sky⟧ is the same in the two uses. It seems that the only difference is the position of the lexeme in the sentence. Facts such as these are problematic for any account based on the lexical categorization of words, an argument put forward by neo-constructionist approaches, such as the Distributed Morphology approach in Marantz (1997), or the approach advocated in Borer (2005). The current position takes the neo-constructionist view that lexemes are not categorized in the lexicon and instead receive their value in the morphological/syntactic component of the grammar. However, unlike traditional neo-constructionist approaches, the current model does not rely on value-specific abstract features to distinguish verbs and nouns: the distinction is instead based on the complexity of the formal expression used to identify the arbitrary denotation (call this the **target**) of a lexeme. Specifically, a lexeme is assumed to be verbal if it identifies its target as a fully audible constituent and nominal if a simple form is sufficient for the identification of its target. Under this account, the expletive *it* in (1a) must appear to express that *rain* is verbal in the context; when the lexeme is nominal in (1b), it only needs to apply to its target as a simple form.[7]

Before arguing in Section 2.3 that endocentric principles applied to audible forms offer a natural tool to capture the distinction, I first show that there is likely a fundamental reason behind the hypothesis that a formal difference is at the center of the verb-noun dichotomy.

---

[3] See Section 3 of Levin & Krejci (2019) for an overview of the literature on this issue.

[4] According to Marchand (1960: 303), most nouns converted from verbs "denote the act or rather a specific instance of what the verbal idea expresses." Precipitation lexemes also share this characteristic even though they are denominal (verbs derived from noun) according to Levin & Krejci (2019).

[5] I borrow the double square bracket ⟦…⟧ notation of formal semantics to informally identify denotation.

[6] Levin & Krejci (2019) position is similar, although stated in different terms. For them, the precipitation noun is the sole participant of the denominal verbal event (no other participants implied) unlike cases of denominal conversion.

[7] This does not mean that nominal *rain* cannot apply as a constituent – it can when it is modified, for example, *(the) cold rain in September*. Simply, being a constituent is not a requirement for the definition of nominal value: a simple form within a DP or a bare form (e.g. *rain is wet*) is enough to identify the target.

2.2 Linguistic form, substance, and space-time

Given the arbitrariness of the sign (de Saussure 1916), the "targeting potential" of any simple formal unit of the lexicon[8] is identical to any other: the form *x* or the form *y* can equally apply to any arbitrary target, as each application is based on the lexical conventions of the language. The external realities (or denotations) that can be arbitrarily identified by formal objects of the lexicon, however, might not all have the same requirements for a proper formal identification. This is illustrated by comparing the formal requirements needed to identify a specific individual reality (say the ⟦Fido⟧ reality) versus the formal requirement needed to identify multiple individual realities that share properties (say, ⟦dog⟧ realities). In the case of a specific individual, the unique form *Fido* is enough to fully identify the unique ⟦dog⟧ reality ⟦Fido⟧. However, to identify multiple individual ⟦dog⟧ realities, for instance ⟦Fido⟧ and ⟦Rover⟧, a unique form *dog* is not sufficient. A proper identification requires multiplication of the form *dog*: it must be pluralized. The plural nominal *dogs*, under this view, is thus the quantification of the singular form *dog* so that each individual ⟦dog⟧ reality can be identified by its own copy of *dog*.[9] In other words, outside of the conventions of English, the formal objects *Fido* and *dog* can apply to any arbitrary reality of the world. The fact that they apply to distinct and specific realities is a question of the lexical conventions of the language, and the realities of what they apply to, given these conventions, impose specific restrictions on their formal realization because *dog* denotes a set, while *Fido* denotes an individual.

The different formal patterns provided by grammar – e.g. the singular bare form *x* and the plural bare form *x*+*s* – effectively provide some information regarding the target of these forms: to capture this intuition, I will say that forms in grammatical contexts have an Application Force (AF). Simple lexemes in the lexicon have no AF because their target is arbitrary and thus dependent on social conventions: the segmental composition of a simple lexeme provides no information about what it applies to. In audible morphological or syntactic contexts, however, the lexeme has an AF: the presence or absence of the *-s* affix creates an audible opposition that correlates with different kinds of targets. A singular (non-quantified) bare form *x* targets a specific individual named *x* (a proper noun); a bare plural (overtly quantified) form *x*+*s* targets a set of individuals that can each be identified with the singular form *x* (plural count noun); the DP *a(n) x* targets a unique member of the set of things that can be identified with *x* (singular count noun); and so on.

The hypothesis is that the nominal and verbal values have different AF because their targets are identified by audibly different expressions. Specifically, the AF of a lexeme in a nominal position corresponds to any external reality for which a unique form is sufficient for proper identification. This is why substantive notions are the typical target of nominals: these are readily identified with a unique form, with or without quantification – *Mary*, *dog*, *rice*, etc. The target of a lexeme in a verbal position, however, must be identified with a complex form (a constituent), because it targets aspects of space-time. This means that at some level, the noun-verb dichotomy is related to the two kinds of realities that our senses can fundamentally recognize: substance (or matter), the solid stuff that makes up our world; and the space-time continuum, where substance (ourselves included) appears. Furthermore, it has long been recognized that humans'

[8] To be clear, *simple formal unit* refers here to lexical forms that have no internal (morphological) structure.

[9] Copying of the singular to express plural is explicit in languages that use reduplication: in Pangasinan, discussed in Rubino (2001), reduplicated *manómanók* 'chickens' is the plural of singular *manók*.

understanding of the space-time distinction is rooted in the materiality of substance.[10] For instance, I have an understanding that my body occupies a certain volume of space, yet I cannot separate this volume from my body: if I move my body, then the volume it occupies moves with it. More generally, for humans to evaluate a gap of space, two distinct material references are needed. Similarly, our understanding of time is also dependent on substance, as it is rooted in changes seen in the material world.[11] Arguably, then, the identification of substance and of space-time require different formal modalities because of their materiality (or lack of materiality). Substance (or material reality) can be formally identified directly with a simple term because it is manifest. In contrast, identifying a "slice" of space-time with a simple term is impossible because this slice has no materiality: there is nothing that the form can apply to. The proper identification of such targets requires a linguistic expression that occupies a wider temporal space than a single label, a full constituent.

Although one might think that a lexeme's position in a constituent would be dependent on its denotation (substantive or time-space), once endocentric principles are used to distinguish verbal and nominal positions, identification is independent of denotation: this is why conversion is possible in the first place.

2.3 Endocentric positions in audible constituents

The audible constituent is a formally complex expression, made up of two formal objects (two overt forms). The verbal and nominal positions are then defined in relation to the endocentric status (headedness) of the formal parts of this audible constituent.

Endocentricity is a fundamental notion of linguistic theory, which has its origin in the pre-generative era of the twentieth century. Simply put, saying that a complex expression is endocentric means that one element in the expression – its head – is more important than the others. Bloomfield (1933) was the first to use endocentricity in syntax and morphology by introducing an opposition between constructions that are endocentric – with a head – and others that are exocentric – without a head. Dependency Grammar (Tesnière 1959) uses the notion of head as the backbone of its syntax, with words relating directly to heads (without mediation by constituents). Heads and constituents are explicitly tied in the phrase structure rules component of early generative grammar (Chomsky 1957) and ultimately, the head and the phrase are mapped as, respectively, the X and XP of X-bar theory (Chomsky 1970). Central to all these approaches is that the head is the nucleus of a complex expression, the locus of its basic properties. In the current approach, endocentricity determines the target of the audible constituent based on the target of its audible parts. A formal constituent is endocentric if it has a head part and a non-head part – called the **term** of the constituent: the target of the head determines the target of the whole constituent, which is restricted by the target of the term. For example, in *orange cat*, the form *cat* is the head of the constituent, so the constituent identifies a ⟦cat⟧ reality, not an ⟦orange⟧ reality. As the term of that constituent, the denotation ⟦orange⟧ restricts the application of *orange cat* so that the constituent not only identifies ⟦cat⟧ realities, but those ⟦cat⟧ realities that can also be identified with *orange*.

Returning to the nominal and verbal values with this in mind, I assume that the following (preliminary) conditions are central for the definition of the verb-noun dichotomy:

[10] This is clearly articulated in Newton's (1689) *Scholium* (thanks to Wayne Myrvold for the reference).

[11] It is impossible to determine whether two identical pictures were taken at different moments in time or if they are copies of the same scene at one moment unless there are some differences (a change) between the two.

(2) a. For a lexeme to be verbal, it must be the tense-inflected head of an audible constituent.[12]
b. For a lexeme to be nominal, its highest projection must appear in the term position of an audible constituent.[13]

The head and term relevant for (2) are determined by a word-order parameter *à la* Principles and Parameters (Chomsky 1981). In English, the default parameter for syntactic constituents is head-initial: the head precedes the term of the constituent. The head-final value is the special case, which, for the current discussion, is restricted to the inflected head of sentential constituents: the inflected head of a sentential constituent follows the term of the sentential constituent.

By defining a verb in (2a) as a form that must be in the position of head of an audible constituent, it follows that a verb will always be audibly realized as a full constituent. As stated above, an audible constituent is always a complex constituent, so alone, a simple form cannot be an audible constituent and thus cannot be the head of a constituent. Consider, for example, the precipitation verb *rain* in (1a) above. The presence of the expletive *it* in *it rains* is needed for the position of *rain* to be a head (and a verb under (2a)): it is an inflected head in final position of the constituent. Its arbitrary denotation ⟦water falling from the sky⟧ is identified as a full constituent. And given that the denotation of *rain* is "complete" – no other participant is required to identify ⟦water falling from the sky⟧ – the term of the head-final sentential constituent can be an expletive, a pronoun that has a formal audible realization (overtly marking the audible constituent), but no referential antecedent. Assuming that the AF of a verb is an aspect of space-time – regardless of denotation – related to tense, then *it rains* identifies that ⟦water falling from the sky⟧ is stretched over a slice of space-time. Being sentential, the constituent *it rains* is a proposition: it identifies a type of situation in the world that can have a truth value.

This analysis of the verbal use of *rain* shows that even if the source of the verb-noun dichotomy can ultimately be traced to requirements regarding the identification of external notions such as space-time and substance, the distinction framed in terms of the endocentric audible constituent is independent of these requirements. To put it differently, denotation and formal identification are not engaged in strict equations of the type ⟦external substance⟧ = "nominal/term position" and ⟦space-time specification⟧ = "verbal/head position." This approach to grammatical values implies an autonomy of formal objects with respect to denotational content (i.e. Semantics), as claimed back in Chomsky (1957). The conversion case of *rain* shows this since the substantive denotation ⟦water falling from the sky⟧ appears in head position of the constituent, which is also the locus of space-time content. The consequence is that the term of the audible constituent is filled by a dummy pronoun (the expletive *it*): although it does not contribute to the substantive content of the proposition, the dummy pronoun must appear at the linear level to provide a formal object in subject position so that denotation of *rain* is identified by a full constituent.

Turning now to *the rain is cold* in (1b), *rain* is nominal under the definition proposed in (2b). Under the DP hypothesis of Abney (1986), *rain* is a lexeme in the term position of the head-initial DP *the rain*. The lexeme *rain* alone is also nominal when it is a term with a mass-noun reading in the head-final sentential constituent *rain is wet*. When *rain* is used as a noun, its target (i.e. the denotation ⟦water falling from the sky⟧) is no longer the head of a proposition. It is a "participant" in whatever the space-time content of the verbal head of the sentence targets. The denotation of

[12] Although prepositions are not discussed here, they would be defined as uninflected heads of audible constituents: they target aspects of space-time without relating to Tense because they are not inflected.
[13] The reference to "highest projection of the lexeme" is needed to allow nouns to head their own constituent.

rain ⟦water falling from the sky⟧ is constant when used as a noun or a verb: if someone utters that *it has been raining for 6 days* followed by *this rain is worrying the farmers*, it is clear that the difference does not come from the contribution of *rain* in the two uses, as it must be the same. Grammar is then a Logic of Labeling: the interpretation of *rain* as the main predicate of a proposition (as a verb) or as a participant (as a noun) of another proposition is not dependent on its denotation ⟦water falling from the sky⟧ but on the position of its form in the audible constituent.

I conclude with one last point that connects to the conversion cases discussed next. Since the expletive *it* has no target (no antecedent and arbitrary denotation), it does not restrict the sentential constituent: as just mentioned, the content of the proposition comes from the denotation of the head. Nothing, however, prevents adding a subject term with a denotation to restrict the head. A case like *a light rain rained on my head* – example (6a) from Levin & Krejci (2019) – illustrates the issue. The situation identified is then more restricted than the one with an expletive *it* because the subject has lexical content: of all possible raining events, only one where the rain is light are targeted. In the cases of conversion discussed next, all verbal uses must have a lexical subject that restrict the application of sentential constituents.

2.4 Extending the analysis

In the conversion cases discussed in chapter V of Marchand (1960), the transparent verbal and nominal conversions – without obvious changes in lexical denotation – are deverbal (verb to noun conversion). Amongst these deverbal nominals, many are like *rain* in that their denotations have a substantive component. For instance, many manner-of-motion verbs like *walk, run, swim, trot,* etc., have a substantive component (information about the materiality of the event),[14] but are arguably more verbal than nominal since they necessarily entail motion.

The denotation of *walk* in (3a), roughly ⟦default movement with legs⟧ illustrate this:

(3) a. *Mary walks.*
b. *a walk*

Unlike the "complete" substantive denotation of *rain*, however, the lexical denotation of these verbal lexemes requires the presence of a lexical subject (an argument in the traditional sense, as is the case for most verbal lexemes of English). This is why it is not possible for the pronoun *it* to have an expletive interpretation in *it walks*: it must have a non-human antecedent. The denotation of these verbs is thus incomplete: although *walk*'s denotation refers to legs, the lexeme does not identify the owner of the legs. Therefore, subject form *Mary* in the sentence must identify the walker in (3a). The requirement that an actual subject be present with the verbal use, however, is only a requirement of the verbal use. It is still perfectly possible to use the lexeme as a nominal. Given that the identification for nominal value requires only one form, an expression like (3b) necessarily implies that information outside the NP provides the content supplied by the subject in the verbal use (3a). For instance, if Mary mentions that she goes for a walk every day, the implication is that the legs involved in the walk are Mary's. Evidently, more formal content can be added to the nominal head (e.g. *Mary's daily walk*) but these additions are not required and, if present, only serve as restrictions on the target of the noun.

Manner-of-motion lexemes, and many other similar kinds with denotation of gestures or an

[14] The assumption that such verbs are different because some of their content is outside of grammar is not new. For instance, Jackendoff (1990) refers to "3D spatial features" to encode lexical content that is not truly linguistic.

audible manifestation of gestures (such as *laugh*, *sniff*, *sneeze*, *whimper*, etc.), tie the temporality of the event they describe to substantive denotation: the substance overtly manifests the space-time content of the lexeme. English's most basic verbs (e.g. *be, stay, go, move,* etc.), however, provide no substantive content at all: their content is only a space-time frame. Consequently, the basic verbal use of such lexemes must appear with subjects the denote substantive content for the overt manifestation the space-time frame. Consider the lexemes *move* and *change* in (4a, b). Unlike *rain* and *walk,* only the subject provides the substantive to identify their space-time frame:

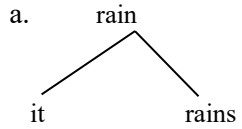
(4) a. *Mary moved.* a.' *The move.*
b. *Mary changed.* b.' *The change.*

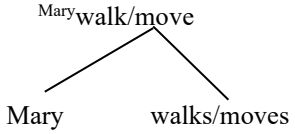
While the substantive subject is the reality "affected" in the space-time frame, the exact consequence is open to many interpretations. For example, (4a) with *move* would be true if Mary changed her position, moved a limb, wobbled a bit, went to another room, changed apartments, etc. The case with *change* in (4b) is open to even more interpretations, as it can apply to many imaginable states of Mary (her hair color, her state of mind, her position, her height, her clothes, etc.). Regardless of this lack of substantive denotation, these items can be used as nominals, as shown in (4a'-b') and the specifics of their interpretation must come from the context. Additionally, even though both the nominal and verbal uses are quite polysemous, it is also easy to create transparent contexts where the use of the lexeme in the two uses denotes the same situation. For instance, imagine that (4a) is interpreted as indicating that Mary has moved to a new apartment. It is then perfectly possible to refer back to this event with the nominal use (4a'), as in *the move was difficult,* to qualify the move. And if *Mary moved skillfully* to avoid an opponent, it would be possible to say *that was a skillful move!* to refer back to the event.

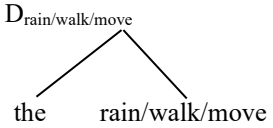
Examples of this kind can be multiplied at will. According to the current approach, this is possible because what is relevant for the verb-noun distinction is tied to the complexity of the formal expression, which is independent of the denotation of lexemes. Under the proposed approach, any lexical form and its target – the arbitrary denotation it is tied to by lexical convention – can appear as a noun or a verb depending on its position in the audible constituent. That a given word can appear either as verb or noun while its denotation is constant is, in fact, not only in the range of what is possible under this account, but to be expected. In Section 3, I discuss how audible form can be used to formalize the approach without referring to any additional abstract feature.

3. Formalizing the audible constituent

Ignoring inflection and the contribution of the determiner, (5a-b) is a preliminary representation for intransitive sentences headed by *rain*, *walk* and *move,*[15] while (5c) is for DPs containing them:

(5) a. 

b. 

c. 

[15] The analysis in (5a-b) is sufficient for the current discussion but would need revisions to take into consideration the hypothesis of Section 3.2 on the nature of verbal content and the tense-inflection head, an issue that will not be tackled in the current paper.

The representation of each constituent (the top of the structure) is made up of copies of the lexical input forms, as well as the label D for the determiner for (5c) in place of its content (not discussed here). The input level is the audible constituent – made up of two input forms [α β] or [the α] at the bottom of the structure. In all cases, lexical input form is copied at the level of the constituent (i.e. $^{\alpha}\beta$, or D_{α}), with the head final *move/walk* in regular script, the subject term Mary in superscript, and the terms of the DP in subscript; the case of verbal *rain* in (5a) differs from the other two verbal cases in (5b) in that only the head-final rain (the β form of the input) is copied at the top.

In Section 3.1, I discuss the rationale behind these representations, and in Section 3.2, defend the hypothesis that, unlike substantive content, which is denotational (outside of grammar), distinctions of space-time belong to the verbal system. Assuming that *be* content is a special verbal element in English, as it is a bare inflectional head that licenses subjects, allows us to provide a clear distinction between the contribution of formal roots in verbal and nominal uses.

3.1 Endocentric alignment rules

The representations in (6) encode information about the relationship between the input forms in two ways: variations in script size and variations in script position. For script sizes, the regular script corresponds to the head element of the constituent, which is verbal for the lexemes in (5b) and the determiner in (5c), and the smaller superscript/subscript size corresponds to the lexemes that are terms and thus nominal. Using differences that are tied to form, like size of script, to represent the different values of the lexemes means that there is no need to stipulate independent features (like V or N) to express these roles in the grammar. Once we have established the convention that, in the representation of a constituent, regular script size corresponds to heads and smaller script size to terms, then with the definitions in (2) we can see that *rain/walk/move* are verbs in (5b) and nouns in (5c). In other words, notions traditionally symbolized by independent features come out of the abstractions built out of the necessary input symbols in (5).

The difference in position with respect to other elements – whether a term is superscripted or subscripted, or not – represents the alignment of input forms. I use alignment to refer to the abstract "stacking" of objects that are otherwise linearized (presented one after the other). In a theory where the primitives are audible forms, as opposed to the abstract features, type, argument structure, etc., of traditional frameworks, it makes sense to assume that constituency is an alignment operation. As Bouchard (2002) points out, audible forms must be linearized as a matter of necessity because they are made of objects (phonemes) that must be articulated one after the other. Phonological segments must be linearized (produced at different moments) and only suprasegmental units (e.g. stress and tone) can be produced at the same time as a phonological segment (specifically vowels). In Optimality Theory, the placement of suprasegmental units relies on Alignment Constraints (see McCarthy & Prince 1993). I borrow this notion of alignment; I hypothesize that forms presented linearly at the input level are abstractly aligned by syntax at the output so that they are a unit – they all apply at the same time. This can be formalized with the rule in (6), where relative position of input form along the horizontal axis (the dotted line) is used to represent alignment:

(6) Alignment rule: $\cdots\alpha+\beta\cdots \rightarrow \ldots{}^{\alpha}\beta\cdots$

Rule (6) indicates that α and β are combined (+) and aligned so that on the right of the arrow, the first (α) appears slightly higher on the horizontal axis, and the second (β) slightly lower. Input forms α and β are audible, but the output, the audible constituent, is an abstract arrangement of copies of input forms. This expresses that α and β apply to a target at the same time, as a unit.

The representations in (5) encode both endocentric values, with the differences in script sizes, and constituency, through differences in alignment. The linear level – the bottom of the structures in (5) and to the left of the arrow in (6) – corresponds to the audible aspects of the constituent; the constituent level – at the top of the structures in (6) and to the right of the arrow in (7) – corresponds to the compositional denotation of the constituent. In the case of *rain* in (5a), the alignment rule (6) does not apply. With precipitation lexemes, the presence of the expletive subject at the linear level provides the formal manifestation that indicates that *rain* is verbal and propositional (an inflected lexeme in head-final position): the target of the constituent only comes from the head. This is a specific labeling pattern of English, which applies to precipitation lexemes. Most intransitive constructions of English rely on a pattern where a nominal term subject provides denotational content for the proposition. Consider the representation for ^{Mary}walk in (5b). Given rule (6), the audible objects *walk* and *Mary* are interpreted as aligned with one another: they apply to a unique reality. Given that *walk* is the head of the audible constituent (inflected head-final and the regular size script in the abstract configuration), the target of *walk* is identified by a fully audible constituent (*Mary walks*). The constituent *Mary walks* thus identifies a situation of walking, which is restricted to one individual, the one identified by the simple form *Mary,* a noun. When these lexemes are part of the DP in (5c), they identify their target as simple forms because they are the terms of the constituent headed by the determiner: the determiner, in a way, "directs" its term complement (nominal lexeme) towards a certain part of the discourse.

3.2 Verbal content and the head of the sentence

The denotations of the constituents in (5) contain the phonological matrices of the lexemes of the input, under the assumption that each of these matrices identifies an external reality according to conventions in place when the form is used. With *Mary*, *rain* and *walk*, the identification is clear thanks to the manifest (substantive) content of ⟦Mary⟧, ⟦water falling from the sky⟧, and ⟦default movement with legs⟧. For the determiner *the* in (5c), it is the function of *the* that is significant, not its form. I argue here that the space-time frames associated with the content of lexemes like *move* and *change* in Section 2.4 is also functional in the sense that it is part of their description.

The assumption is that the lexical content of verbs is a characterization of space-time frames because such frames cannot be defined without substantial content, as discussed in Section 2.2. It is not possible to isolate space-time frames in the world so that they could be identified with a label. To illustrate, consider the contrasts in (7), made up of three verbal lexemes *is*, *arrives* and *goes,* that have no substantive content the way *rain* and *walk* do. There is no materiality of substantive denotation to provide an overt manifestation of their space-time frames, which are only understandable through the denotations of *Josie* and *inside*.

(7) a. *Josie is inside.*
b. *Josie arrives inside.*
c. *Josie goes inside*.

The different space-time frames can be understood in term of spatial alignment between ⟦Josie⟧ and ⟦inside⟧: *is* in (7a) implies that ⟦Josie⟧ and ⟦inside⟧ are spatially aligned; with (7b-c), the situation implies no spatial alignment at the start, the understanding being that spatial alignment is imminent with *arrive* (7b) but requires more time for the case of *go* in (7c). It is the relationship of the substantive elements to each other that makes the different space-time frames visible. If the three situation types described in (7) are stripped of the denotations of ⟦Josie⟧ and ⟦inside⟧, the

verbal contribution just stated has no reality.

Since the contribution of verbal content cannot be tied to any denotational content, it must be part of the verbal system of grammar. Each of the verbs in (7) have a specific space-time frame contribution, which accounts for the different situations these expressions apply to in (7). A full discussion of this content is beyond the scope of this paper, but I nevertheless want to touch on the basic verbal space-time unit of grammar, which I hypothesize is formally realized by the suppletive paradigm of *be* in English. When this space-time unit is aligned with one uninflected (singular) term, the result identifies an individual. In (7a), then, the formal part of the expression that identifies an individual is not the formal object *Josie*, but rather the constituent *Josie is*.[16] The hypothesis is that *be*'s contribution is unique in the grammar, as it is the only object in English that licenses a subject. To put it differently, *be* is a bare inflected head and is responsible for the presence of subjects in all tensed propositions of English. All other verbs of English are roots, which, in simple tenses, are placed directly where the suppletive forms *is/am/are/was/were* would appear in a linear sequence. That place is the head of the proposition. In intransitive propositions, the content of verbal roots like *stay*, *arrive*, *go*, etc., adds space-time content to the basic frame provided by *be*; roots like *walk*, *run*, *swim*, etc., also add a denotational content, as would any nominal lexemes in conversion cases. When the root's denotation provides all the denotational content needed to identify the situation (e.g. *rain*), the denotational content from the subject licensed by *be* is not needed. Unlike propositions with a lexical subject, in which the whole situation is delimited by the entity identified by the subject, the denotation ⟦water falling from the sky⟧ is not delimited by a subject. When *rain* is head of the proposition, its denotation is spread over the whole space-time frame provided by the constituent.

4. Conclusion

I have used data about transparent verb-noun conversion to argue that fundamental aspects of linguistic analysis can receive a principled account that only makes reference to the audible parts of linguistic expressions. The underlying separation between linguistic form and denotation in this model provides a direct account of why many lexemes can freely alternate between nominal and verbal values without their denotation being affected: a lexeme is verbal, or nominal, given where its formal matrix appears in an audible constituent. This account requires no grammatical abstractions (zero-morphemes or *a priori* category features) or non-grammatical notions (discourse functions, denotational content, semantic/thematic roles): in other words, when syntax operates on linguistic form, it is autonomous, as anticipated in Chomsky (1957). I also proposed that *be* provides the basic space-time frame that identifies sentences and licenses subjects in English: verbal roots of English are placed onto this basic frame. They add a space-time or denotational content to the constituent. Their contribution is nevertheless independent of that of the inflected head of sentences. From a formal standpoint, this is why they are freely allowed to appear as term with nominal value in the audible constituent.

[16] That *Josie is* means 'existence of Josie' follows from the composition of the expression: the basic time-space unit (*is*) is aligned with bare form *Josie*. This constituent identifies an individual (*Josie*) and the space-time they occupy (*is*) at the moment of speech. Application of *Josie is* implies a situation where Josie exists.

Acknowledgements

I wish to thank Juhani Dickinson and Mark Daley, who provided attentive ears to the central ideas presented, as well as to the participants of the 2024 LACUS annual conference, especially Eric Wheeler for his enthusiastic response after my presentation. And I particularly want to thank an anonymous reviewer who made me realize that the focus of the original written version was all wrong. The reviewer's comment forced me to completely change the paper, and I think for the better. And, as always, thanks to Jennifer O, for everything.

References

Abney, Steven Paul. 1987. *The English noun phrase in its sentential aspect*. Massachusetts Institute of Technology. (Doctoral dissertation.)

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Borer, Hagit. 2005. *Structuring sense volume I: In name only*. Oxford: Oxford University Press.

Bouchard, Denis. 2002. *Adjectives, number and interfaces*. Amsterdam: Elsevier.

Chomsky, Noam. 1957. *Syntactic structure*. The Hague: Mouton.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick, Jacobs & Rosenbaum, Peter (eds.), *Readings in English transformational grammar*, 184-221. Waltham, Mass: Ginn.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

de Saussure, Ferdinand. 1916. *Cours de linguistique générale*, 1st edn. Preface by Bally, Charles & Sechehaye, Albert (eds.), Paris: Payot.

Fodor, Jerry. 1998. *Concepts: where cognitive science went wrong.* Oxford: Oxford University Press.

Jackendoff, Ray. 1990. *Semantic structure*. Cambridge, Mass.: MIT Press.

Levin, Beth. & Krejci, Bonnie. 2019. Talking about the weather: Two construals of precipitation events in English. *Glossa: a journal of general linguistics* 4(1). Article 58. 1-29.

Lieber, Rochelle. 2005. English word-formation processes. In Stekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-Formation*, 375-427. Dordrecht: Springer.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Dimitriadis, Alexis & Siegel, Laura & Surek-Clark, Clarissa & Williams, Alexander (eds.), *University of Pennsylvania working papers in linguistics* 4(2). 201-225. Philadelphia: University of Pennsylvania Working Papers in Linguistics.

Marchand, Hans. 1960. *The categories and types of present-day English word-formation*. Wiesban: Otto Harrassowitz.

McCarthy, John J. & Prince, Alan. 1993. Generalized alignment. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of morphology,* 79-153. Dordrecht: Kluwer.

Newton, Issac. 1689. Scholium to the definitions. *Philosophiae naturalis principia mathematica*, Bk. 1. (Translated by Motte, Andrew (1729), revised by Cajori, Florian (1934)), 6-12. Berkeley: University of California Press. https://plato.stanford.edu/entries/newton-stm/scholium.html

Rubino, Carl. 2001. Pangasinan. In Garry, Jane & Rubino, Carl (eds.), *Encyclopedia of the World's Languages: Past and Present*, 539-542. New York/Dublin: H.W. Wilson Press.

Tesnière, Lucien. 1959. *Éléments de syntaxe structurale.* Paris: Kincksieck.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:3.28-39 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative CommonsAttribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

The voice of the unheard: A linguo-cultural worldview of Canada in the Polish immigrants' memoirs of the early 20th century

Maciej Laskowski
Adam Mickiewicz University

Abstract: Similar to other parts of North America, Polish immigration to Canada occurred principally throughout the 19th and 20th centuries, with the earliest major wave in 1858. Some were motivated by the country's opportunities in fur trading, while others sought refuge from conflict and persecution. All left their homeland to forge a brighter future for themselves and their descendants. Today, the small yet vigorous communities of Poles in Canada not only add to this country's cultural diversity but also form a meaningful part of its history and identity. For this reason, the displacement encountered by the Polish diaspora necessitates a specific analytical framework situated within the field of language and culture relationships between Poland and Canada. One key aspect of this displacement is the linguo-cultural worldview of Canada documented in the early 20th century memoirs of Polish people, an analysis of which constitutes the study's primary goal.

Keywords: linguo-cultural worldview of Canada, the voice of the unheard, Polish migrants' memoirs, the early 20th century

Languages: English, Polish

1. Introduction

Migration – understood as the movement of people from one location to another with the intention to settle permanently or temporarily and over various stretches of land (Merriam-Webster 1993: 1432) – has been occurring globally since the beginning of humankind and for a number of reasons. An example of such a process is the Polish migration to Canada, documented in various types of memoirs. "Memoir has turned out to be a momentous genre in Canadian literature throughout the centuries as the genre has enabled different previously unrepresentative groups such as migrants and members of ethnic minorities to find their voice and express their experiences." (Drewniak 2023: 1) Such primary sources are thematized and well represented but mainly those written within the last 100 years, particularly in the span from after World War II to nowadays. One study fulfilling the above criteria is Melchior Wańkowicz's *Tworzywo* (1954), translated in 1973 by Krystyna Cękalska as *Three generations*. Other examples are Jacek Kozak & Joanna Lustanski's *Zwyczajna przeprowadzka: Pamiętniki imigrantów z Polski do Kanady (1988–2012)* [An ordinary move: Memoirs of Polish immigrants to Canada (1988–2012)] from 2014 and the article *From*

nostalgia to locus amoenus: Polish migrants' memoirs in Canada and the idea of home, identity, and belonging (2023) authored by Dagmara Drewniak. However, critical studies on memoirs from the early 20th century are evidently few and far between. This is unsurprising, as fewer memoirs from earlier periods have survived.[1] Memoirs of Polish migration to Canada from this period are dealt with by Krystyna Romaniszyn in *Obraz Polski a świadomość narodowa: Analiza pamiętników imigrantów polskich w Kanadzie* [The worldview of Poland and national consciousness: An analysis of the memoirs of Polish immigrants in Canada][2] from 1986. As the title suggests, the main focus of Romaniszyn's article is the worldview of Poland, however an analogous linguo-cultural worldview of Canada, written in English, is virtually non-existent.

Therefore, the main purpose of the following article is to present the linguo-cultural worldview of Canada (LWC) perpetuated in selected memoirs of Polish immigrants in the early 20th century, which, at the same time, "started the Polish memoir writing in Canada" (Śliz 2009: 47) and "have constituted until the present-day the documentation of the great peasant epic of the beginnings of the 20th century." As implied in the title, this work concerns *the voice of the unheard*; the selected memoirs are those written by the have-nots of peasant origin: uneducated petty farmers and unskilled laborers, all of whom were first-generation immigrants. The discussion also provides their concise profiles. In the end, it is hoped that this paper will help to honor and foster the memory of the Polish authors and their writings and save them from oblivion.

2. Outline of Polish migration to Canada

The permanent, uncontested presence of Poles in Canada began at the end of the Seven Years' War (1756-1763), when Polish soldiers served as mercenaries in German units within the victorious English army. More Poles arrived in Canada from the United States after the American War of Independence ended in 1783, alongside numerous Anglo-Saxon loyalists (Zins 1975: 253). Shortly after the First Partition of Poland (1772), other groups of Polish immigrants arrived in Canada. Significant waves occurred later, between 1854-1901, 1902-39, 1944-56, 1957-79, and 1980-93,[3] with the first wave leading to the establishment of the first Polish town, Wilno, in Canada in 1858 (Heydenkorn 2019).

The first immigration waves were triggered by the loss of homeland, religious persecution, dire straits in overpopulated Polish villages, and economic stagnation in urban areas. When in Canada, "lack of education, illiteracy, rural lifestyle, attachment to the church and traditions effectively impeded the process of integration of Poles into the mainstream of social and cultural life." (Śliz 2009: 45, our translation) This period saw an even larger wave of Ukrainian emigration to Canada, primarily from the lands captured by the Austrians (Janowska 1971: 20).

A notable shift occurred during the third *significant wave*:

> While the majority of the first Polish newcomers to Canada did not arrive directly from Poland, the immigrants of 1916 to 1939 arrived from an already independent Poland and settled […] in Manitoba, Saskatchewan, and Alberta, where they built farms or worked in railway construction or coal mines. The second generation of

[1] In fact, the early Polish emigration to Canada produced prolific oral tradition, which was transplanted to Canada. Unluckily, due to a lack of Polish press there, this process was hindered (Ożóg 1984: 41 after Sojka 2010: 292).

[2] The English translation of the title is ours.

[3] Some, for example, Anna Śliz (2009: 47), propose three such periods: the turn of the 19th and 20th centuries, the after WWII wave, and the post-Solidarity emigration of the 1980s and 1990s.

> Polish immigrants often moved to larger settlements or towns, where they opened small businesses. (Heydenkorn 2019)

The post-WWII wave included former soldiers, concentration camp survivors, and political refugees from communist Poland. Half of these immigrants settled in Ontario. This post-WWII group, along with the second generation, was more financially secure and had professional experience and technical qualifications. They established their businesses, held leadership positions in industry, and were notable in social services, healthcare, and higher education (Heydenkorn 2019).

During the era of real socialism in Poland, emigration levels to Canada fluctuated. There were times of strict restrictions on leaving the country, alternating with periods of relaxation (the Polish thaw). This pattern continued until the 1980s when political changes in Poland prompted another wave of Polish emigration to Canada (Śliz 2009: 46).

Today, approximately one million Canadians identify as having full or partial Polish heritage, with nearly 200000 Canadians speaking Polish as their native language (Heydenkorn 2019).

3. Material and methodology

The research material is composed of four memoirs written by Polish immigrants in Canada, selected from a total of 33[4] that were collected for the contest held among *Polonia* by the Institute of Social Economy (*Instytut Gospodarstwa Społecznego*) in Warsaw in 1936. Due to the anonymity of the authors, they will be referred to as Memoirist 3, Memoirist 8, Memoirist 10, and Memoirist 16 respectively, which corresponds to the way they are alluded to in the collection of the memoirs submitted for the contest entitled *Pamiętniki emigrantów: Kanada, nr 1–16* [Memoirs of emigrants: Canada, 1–16][5] edited by B. Ratyńska in 1971.

The analyzed primary texts are invaluable for the presentation of LWC. Even though the memoirists' main task for the contest was to sketch an image of Poland, having experiences in Canada, they did it against the backdrop of that country. In this way, the memoirs are an indirect source of learning about Canadian culture. Subjective and imprecise though their recordings may be, the perspective of a particular memoirist rooted in a specific time and place helps the reader to grasp the reality then, more fully and complement what history books offer. Furthermore, in the preface to the memoirs' collection, Andrzejewski (1971: 14) writes that dealing with the memoirs "is not just a matter of reverence for the achievements of the past" but of "a living book." Ultimately, the sheer fact that the Polish peasants (of the time) formed an enormous majority of the nation makes it worthwhile to handle their views.

The criteria for the chosen texts primarily revolve around the common features of their authors (see Section 5).

The method used in this study is qualitative, based on the extraction and content analysis of language samples contained in the research material, imbedded in immigrant writing. The analysis focuses on identifying key themes such as culture, economy, politics, law, administration, and the environment that illustrate the immigrants' worldview. Because its definitions are multiple, contested, and not finite, in the subsequent paper, immigrant writing is understood as recordings characterized by reportage features that have

[4] 16 out of 33 were published.
[5] The English translation of the title is ours.

> […] basis in real-life events and the author's first-hand experiences […]; references to multiple cultures and/or geographic locations […]; common themes and motifs (for example, locality versus globalism, community versus nation, exploration of belonging and identity issues, comparison of the home and host countries and cultures, living in a multicultural community, intercultural relationships, creation of new traditions and heritage, revision of gender roles and the like). (Kosmalska 2022: 345)

To present the linguo-cultural worldview of Canada as expressed by specific individuals (Polish immigrants arriving in Canada and describing it), we need to look at two complementary terms: *linguistic worldview* and *cultural worldview*. The first takes from the definition of Jerzy Bartmiński (1999: 104), who sees it as

> […] an interpretation of reality contained in language, which can be expressed in the form of a set of judgments about the world. These may be judgments either recorded in the language itself, in its grammatical forms, vocabulary, clichéd texts (e.g. proverbs), or implied by the forms and texts of language. (our translation)

The underlying concept of *cultural worldview* – proposed by Toelken (1996) – is based on the assumption that objectivity of a worldview

> […] actually varies widely according to the viewer's means of perceiving it. Often those means are affected by cultural and linguistic factors that have been so deeply engrained in the mind that they have become a method of thinking which includes a logical system and a set of evaluative assumptions. (p. 263)

4. Timeframe

The logic behind the timeframe lies, again, in honoring "the unheard." In consequence, it covers essentially the lifetimes of immigrants who arrived in Canada between 1902 and 1915, that is, during the second of the "six significant waves of Polish immigration to Canada." (Heydenkorn 2019) There were over 50000 Polish immigrants there at the time. Still, the numbers cannot be exact, as, not infrequently, due to the Partitions of Poland, Poles were classified as Austrians, Russians, Germans, Galicians, Polish Austrians, etc. (Janowska 1971: 19-20). Aside from the already mentioned fact that this period is neglected in terms of analyses of LWC illustrated in memoirs, it was also harder for Poles to settle down in Canada in this period. Firstly, they still had no homeland, for the partitions ended only after WWI, and secondly, they were somewhat pioneers heading into the unknown, whose reception by the Canadians was colder than that of the subsequent waves (Janowska 1971: 22). It was they who – after establishing themselves in the new land – brought their family and friends, enabling them to have a better start in the land of the maple leaf.

5. Profiles of the authors

The profiles of the authors align with the criteria for selecting these authors and their writings, which concentrate on the common features of the memoirists, and which could be labeled "the unheard." Thus, the memoirists:

- were born in the 19th century and immigrated to Canada between the years 1902 and 1915
- were first-generation immigrants
- came from poorly educated, impoverished, rural social strata in economically backward regions of South/Southeast Poland
- were employed as farmers or unskilled workers
- had no command of English or French
- were hardworking
- wrote authentically
- sought to improve their financial situation and escape "exploitation and abuse by landowners, clergy and administration workers, not only by the partition administration, but regrettably by the administration of the Polish state as well" (Ratyńska 1971: 493)

Among the memoirists 3, 8, 10, and 16, the only exception to the above benchmark is Memoirist 16 – who is the sole female. She left for Canada in 1930 to join her husband, who had arrived there two years earlier. Although she did not have any paid employment and emigrated much later than the other memoirists, her perception of Canada is very important for *the voice of the unheard*, because the life of women in Poland at the time under study was in most cases very arduous or simply cruel.[6]

Below are a few facts about the writers' backgrounds:

- Memoirist 3: unskilled worker in Ontario, son of a farmhand from around Radom, born in 1881
- Memoirist 8: unskilled worker in Manitoba, son of a small farmer from around Lwów (today's Ukraine), born in 1877
- Memoirist 10: farmer in Saskatchewan, son of a farmer from around Kraków, born in 1886
- Memoirist 16: wife of a hosier, daughter of a farmer from around Kraków, born in 1902

5.1 Language of the authors

It is important to note that original linguistic samples in Polish are not provided in the article due to length constraints. Instead, the article offers authorial English translations of these samples. These translations are not always literal and do not reflect the language imperfections present in the memoirs. These imperfections arise primarily from two factors. Firstly, all the writers originated from underprivileged, rural, and economically disadvantaged regions of Southern and Southeastern Poland, where formal education was often deprioritized in favor of hard physical labor. Secondly, the memoirs were written over a century ago, meaning many expressions and linguistic structures appear outdated or nonstandard by contemporary Polish language norms.

For example, in one passage about Poland, we find: *tam moja ziemia, tam moje rodzice, pradziadowie leżą* 'there lie my land, my parents, and grandparents.' The pronoun *moje* 'my' before *rodzice* 'parents' is no longer considered correct in standard Polish; instead, the correct form is *moi* 'my.' However, *moje* is still used regionally in certain areas of Poland. As such, a literal English translation of this Polish phrase might read: 'there lie my land, me parents, and grandparents,' illustrating both the grammatical irregularity and the regional variation in the

[6] The appalling situation of women in Poland across the centuries is depicted, for example, in Anna Kowalczyk's *Brakująca połowa dziejów: Krótka historia kobiet na ziemiach polskich* [The missing half of history: A brief history of women in the Polish lands], our translation (2018).

original text.

6. Analysis

The primary factor determining the criteria for the object of study – that is, the aspects constituting the linguo-cultural worldview of Canada – is the thematic focus typical of Polish economic emigrants during the period under consideration. Traditionally, settlers were primarily concerned, both before leaving their homeland and after arriving in the new land, with survival-related matters. As a result, their writings often focused on the culture, economy, politics, law, administration, and environment of their new homeland. Due to the limitations of this article, however, only four aspects are addressed – those shared by all the memoirists included in the study. These are: socio-cultural, political-administrative, economic, and environmental-architectural aspects.[7]

The texts have been analyzed to reveal the Polish immigrants' worldview of Canada. To achieve this, it was essential to answer critical questions such as: Why did the author write this memoir, and what is its central message? How does the memoir contribute to our understanding of the historical period or cultural context it describes? and How does the memoirist construct their identity and present their life story?

6.1 The socio-cultural aspect

The socio-cultural aspect in the following publication encompasses, among others, the migrants' interpersonal relationships with Canadians, experiences of alienation, command of English/French, as well as Canadians' attitude toward the newly arrived, their treatment, and social distinctions. For instance, Memoirist 8, on his journey to Brandon, Manitoba, praises the manners of the Canadian Pacific Railway workers:

(1) 'As for the railroad service, it can be admitted that we have experienced complete courtesy and respect.' (Ratyńska 1971: 219)

Memoirist 10 has a similarly commendable view of the train service but focuses more on the comfort of travel. Heading to Selkirk in Manitoba, he states:

(2) 'The trains were comfortable, suitable for long journeys, the seats are made into beds [...] so that all passengers can sleep in a bed, as long as they have pillows and a cover.' (Ratyńska 1971: 285)

Earlier, while at the port of Quebec, this person notices a warm reception and problem-free medical attention received from officials during the entry procedures:

(3) 'Generally, newcomers were welcome, there were no visits from any doctors, everyone was healthy whoever could go about their own power.' (Ratyńska 1971: 285)

In turn, Memoirist 16 – the wife of the hosier – appreciates the Canadian theft-free environment:

[7] The division is not very strict, as many aspects entwine, and a given sample might be allocated to various aspects.

(4) '[…] at first, after arrival, I couldn't break the habit of holding my purse tight and watching it carefully like in the country.[8] Here you can carry your money with no worry. Same with different things, materials, food, which stand in front of shops. Nobody watches them, because nobody steals.' (Ratyńska 1971: 461-462)

The same author's outlook can be used to herald unfavorable remarks about the socio-cultural facet of living in Canada. The first fragment refers to some sentimental content displayed at the local cinema, the tickets for which are beyond the Polish couple's financial capacity. The other one is about social anxiety resulting from poor linguistic competence. Here, the writer also presupposes being looked down on or at least sized up by the locals. The question arises whether it is the Canadians who judge the book by its cover or the "Polone" who projects her biased ideas on her neighbors:

(5) 'We don't go to the cinema, because both my husband and I don't like that they show stars wooing. […] For cinema you need money, and looking at beautiful houses in the cinema and then coming back home and living in a poorly furnished hovel is what we don't want.' (Ratyńska 1971: 464)

(6) 'I didn't go out into the city alone because I thought someone might ask me something, and I wouldn't know how to respond. I even tried not to go out on the balcony because I knew that everyone was watching how this Polone looked.' (Ratyńska 1971: 461)

6.2 The political-administrative aspect

This aspect encompasses immigrants' perceptions of the judicial system, public institutions, and democratic principles, with a focus on their effectiveness and the attitude of clerks toward (immigrant) petitioners.

After receiving no payment for five weeks of labor during the harvest in Winnipeg, Manitoba, Memoirist 8, along with several other workers, chooses to bring the issue to the local court. The case is handled promptly and successfully:

(7) 'The judge [...] decided to our advantage and he[9] had to pay the court costs and to each of us his due [...]. This was followed by the payment.' (Ratyńska 1971: 222)

Memoirist 16, residing in a city she refers to as "S," confirms the efficiency of Canadian public institutions and is fond of the clerks' and officials' readiness to serve petitioners. Once again, she contrasts this situation with that in Poland, leaving no doubt as to which she views more favorably:

(8) 'You can get anywhere and take care of things because everyone you asked for information showed you and took you where you needed to go, not like in the country, where any random official would insult you, calling you cattle and boors.' (Ratyńska 1971: 460)

A passage from Memoirist 3, who settled in Ontario, further suggests that Canadian citizenship

[8] Reference to Poland.
[9] Reference to the employer.

officials are both efficient and effective in their operations:

(9) 'I took citizenship of that country in the year 1924 without any difficulty.' (Ratyńska 1971: 120-121)

However, not everything is ideal; the memoirists' accounts are tinged with a sense of disillusionment regarding the attitudes of those in high-profile positions – including administrative workers, local politicians, medical professionals, and affluent individuals – toward immigrants. At one point, Memoirist 3 expresses doubt about the integrity of such individuals, implying that, ultimately, the wealthy will inevitably exploit the poor:

(10) 'And as for the local authorities, they are polite, they serve a person without any delay, right away, because here is democratism [...], here whether a doctor, or a pharmacist, or an engineer or even the biggest bourgeois, they will talk to you as equals. He knows that you are inferior anyway, and he will take advantage of you at every step, because you are unorganized.' (Ratyńska 1971: 119)

6.3 The economic aspect

This section highlights some of the most significant factors influencing the decision to emigrate, including wages, (un)employment, working conditions, the standard of living, the costs of transportation, clothing, food, agriculture, and taxes. Here, Memoirist 8 will be given the primary voice, allowing readers to see how personal opinions are shaped by the prevailing circumstances. In this case, chronology plays an important role. The first three excerpts, all from Brandon, Manitoba, reflect the memoirist's satisfaction with and optimism about his financial situation:

(11) 'It can be said openly and emphatically that at that time there was prosperity if a person only could and wanted to work, and did not waste a dollar on triviality.' (Ratyńska 1971: 223)

(12) '[In 1909,] after working for a year, I was able in those days to pay off all debts, support myself, and secure myself for the winter in case of unemployment. [...] Everyone felt cheerful, content, and did not worry about tomorrow or the next winter.' (Ratyńska 1971: 225)

(13) '[In 1923,] prosperity and recovery were [...] invariably the same as before the war [...]. Farmers would go so far as to ask the peasant for harvest and pay $5-6 a day for ten-hour labor; a man desired to live under such conditions because the job was looking for a laborer. [...] I was very happy that I got to this paradisial and not knowing poverty and misery country for the third time.' (Ratyńska 1971: 235).

The final quotation, once again, albeit subtly, contrasts life in Poland with that in Canada, with the latter being implied as more favorable, even described as a "paradisial." However, as previously mentioned, when the financial situation worsens, so too does the tone of the memoirist's reflections. The following accounts shift from an initial focus on hunger-driven desperate actions to profound pain and distress at being unable to bid farewell to a deceased mother, culminating in a disillusionment with humanity. These observations, too, are from the Brandon area:

(14) ‘So it was 1930 or 1931, and I was so desperate, with absolutely nothing to feed myself. [...] We decided that we were going to repeat our march to the mayor of the city every day, demanding help to live. In the end, we were granted one loaf of bread per day by that mayor.’ (Ratyńska 1971: 243-244)

(15) ‘[In 1934,] Old Mother died without covering funeral expenses. I, in turn, helpless [...] wrote back with pain in my heart and with tears in my eyes [...] that I can neither rescue nor come because I have no money.’ (Ratyńska 1971: 245)

(16) ‘[In 1935,] in these times of ongoing depression, a person must be strongly wary and cautious of whom he is dealing with, because no one can be trusted nowadays because of wickedness, insecurity, injustice, abuses of various kinds and exploitation everywhere you go.’ (Ratyńska 1971: 246)

A similar, though perhaps a less dramatic stance, is adopted by Memoirist 3, who shares his experience from Toronto, Ontario:

(17) ‘[In 1926,] all the savings were taken by the doctors, because what I earned on the land was not enough for me alone, not to mention my wife and four children, who I left in the city. With me I could not take them, because they were still small, they could not help me on the land.’ (Ratyńska 1971: 116)

6.4 The environmental-architectural aspect

The final aspect underlined in the article refers to the Polish immigrants’ perceptions of Canada’s natural environment, including its forests, animals, weather, towns, and houses. These reflections often capture their emotions, particularly during their first encounters with these elements. Unlike in the case of the previous aspects, it is notably more difficult to find negativity in these descriptions. The following example captures Memoirist 8’s initial impressions of Brandon, Manitoba:

(18) ‘Living conditions here were commendable, as the city of Brandon began to expand rapidly: sewerage, electrification, telegraph and telephone wires, and railway in all directions. [...] In a word, prosperity on a large scale.’ (Ratyńska 1971: 225)

As he journeys from Manitoba to Vancouver in British Columbia, the author expresses admiration for the natural landscapes he observes from the train, before offering insights into the city itself:

(19) ‘When you pass Calgary, there are mountains, every one of them higher and overgrown with pine trees, and then cedar trees and lovely waterfalls from these mountains, which form streams below, then rivers smaller and larger [...].’ (Ratyńska 1971: 237)

(20) ‘I was so enthralled by the charm of this area as if enchanted [...]. Here and there one can hear the rasp, swish, friction, some unusual rumbling of ships. The whole city is shaking from this liveliness. [...] I see automobiles splitting in abundance in different directions. [...] And why didn’t I come here earlier?’ (Ratyńska 1971: 238)

Although infrequent, critical observations concerning this aspect do emerge occasionally. One such instance arises from Memoirist 10's journey from Yorkton to Insinger in Saskatchewan. Burdened with carts filled with food and clothing for several families, the horses are unable to pull the weight, forcing the travelers to march on foot:

(21) 'It was a hard road in those days; it took a lot of hardship to travel 35 miles, as one had to walk on muddy roads. Riding on a cart was out of the question, as the load they took with them was already enough weight for the horses. And when we arrived at the creek, we had to carry flour and various luggage on our backs, and several times each.' (Ratyńska 1971: 285–286)

The analysis demonstrates that for every positive comment regarding Polish immigrants' lives in Canada, there is a corresponding negative observation. However, a careful assessment of the memoirists' expressions reveals that the latter are significantly less frequent. As noted earlier, these comments often depend on the specific circumstances faced by each writer at a given time. This observation aligns with Agnew's assertion (2005: 20): "Memories of the marginalized and subjugated illuminate the past in ways that contest hegemonic memories and raise the question of who remembers, why they remember, and what they remember." It is also important to acknowledge that the linguistic examples analyzed represent only a small sample of a much larger body of work. As Drewniak (2018: 151) observes, "the literature of Polish writers in Canada is displaced and appears as a fragmented archipelago of interesting texts and authors."

7. Conclusion

The analysis of the sampled material reveals that the linguo-cultural worldview of Canada, as expressed in the Polish memoirs, is predominantly favorable, especially when compared to that of Poland. Polish immigrants perceive life in Canada as challenging yet easier, fairer, and generally more advantageous than in their homeland. While their opinions and depictions occasionally vary, most of the 16 memoirists whose works have been published (see Footnote 4 in Section 3) express a reluctance to return permanently to Poland. This sentiment is particularly striking considering that all of them harbored a deep yearning for their homeland, families, friends, and an idealized past, despite having left due to harsh and often brutal living conditions.

It is worth noting that although the last-mentioned author (Memoirist 16) experienced Canada long after WWI (i.e. several decades after her predecessors), the LWC outlined by her is not fundamentally different from the image presented by the other authors. This similarity can be attributed, at least in part, to the comparable social status of the writers, all of whom were poor, uneducated, first-generation farmers or unskilled workers. Their shared experiences of displacement and efforts to ascend the social ladder in Canada contribute to this commonality.

The migrants appreciate the locals' unbiased attitudes, honesty, and relatively fluid social classes. They acknowledge the effectiveness of Canadian courts and public institutions, and the availability of modern, affordable transportation. Economic conditions, such as employment opportunities, an improving standard of living, and advanced agricultural practices are also frequently noted. Canada is regarded as a land of opportunity, freedom, and natural beauty. Additionally, experts indicate that the positive impressions of the linguo-cultural worldview were reinforced by the Canadian government's liberal nationality policy, which sought to integrate various ethnic groups. This "Canadian mosaic" provided communities with the space to preserve

their cultural identities (Janowska 1971: 30). As a general rule, there was little anti-Canadian sentiment among the immigrants, contrasting with the anti-Ukrainian sentiments that sometimes surfaced (Ratyńska 1971: 495).

On the other hand, some individuals, like Memoirist 8, did not achieve success: "afflicted with sickness and embittered, they did not even have the money to return home." (Ratyńska 1971: 494) It is important to acknowledge that the writers also encountered cultural challenges, including discrimination, isolation, feelings of inferiority compared to Canadians, and a sense of belonging neither to Canada nor to Poland. The memoirs frequently contain resentful remarks about the demands of hard work, fierce competition in the job market, and unfair treatment from employers. The Polish immigrants often perceived themselves as cheap labor, facing the risk of being the first to be laid off and sometimes not receiving their due wages. Their limited proficiency in English or French further compounded their negative perceptions of the new country.[10] Yet, as earlier indicated, for the majority of the immigrants – both those included in the study and those not included – the new reality in Canada was preferable to their former lives in Poland.

Ultimately, even though the discussed memoirs 1) primarily focus on the worldview of Poland, with the depiction of Canada serving as a secondary outcome, 2) are inherently subjective due to the nature of such texts, 3) are "characterized by a certain severity of notation and simplicity of language" (Drewniak 2023: 1), 4) do not represent the entirety of Canadian regions, and 5) are limited in number and do not fully encompass the entire immigrant experience, they nevertheless: a) capture a portion of the shared history and worldview of the memoirists relating to Canada, b) provide insights into the cultural connections between Poland and Canada, c) "are a meaningful document, which, bringing us closer to the past, at the same time allows a better understanding of today" (Andrzejewski 1971: 15, our translation), d) and give, if only a little, a voice to the voiceless.

References

Agnew, Vijay. 2005. *Diaspora, memory, and identity: A search for home*. Toronto: University of Toronto Press.

Andrzejewski, Adam. 1971. Przedmowa [Preface]. In Ratyńska, Barbara (ed.), *Pamiętniki emigrantów: Kanada, nr 1-16* [Memoirs of emigrants: Canada, 1-16], 5-15. Warszawa: Książka i Wiedza.

Bartmiński, Jerzy. 1999. Punkt widzenia, perspektywa, językowy obraz świata [Point of view, perspective, linguistic worldview]. In Bartmiński, Jerzy (ed.), *Językowy obraz świata* [Linguistic worldview], 103-120. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Cękalska, Krystyna (trans.). 1973. *Three Generations*. Toronto: Canadian-Polish Research Institute.

Drewniak, Dagmara. 2018. "[They] would say she was betraying Poland already": Major themes in contemporary Canadian literature by writers of Polish origins. *Anglica: An International Journal of English Studies* 27(1). 149-165.

[10] In the time handled in this article, a small minority of Polish migrants in Canada spoke one of the two languages. In 1931, over 80% spoke English and in 1941, over 96% (Makowski 1967: 177-178).

Drewniak, Dagmara. 2023. From nostalgia to *locus amoenus*: Polish migrants' memoirs in Canada and the idea of home, identity, and belonging. *European Journal for American Studies* 18(4). 1-15.

Heydenkorn, Benedykt. 2019. *Polish Canadians. The Canadian encyclopedia*. (https://www.thecanadianencyclopedia.ca/en/article/poles) (Accessed June 6, 2024)

Janowska, Halina. 1971. Wychodźstwo polskie w Kanadzie [Polish emigration to Canada]. In Ratyńska, Barbara (ed.), *Pamiętniki emigrantów: Kanada, nr 1–16* [Memoirs of emigrants: Canada, 1-16], 16-30. Warszawa: Książka i Wiedza.

Kosmalska, Joanna. 2022. Defining migration writing. *Journal of Literary Theory* 16(2). 331-350.

Kozak, Jacek & Lustanski, Joanna. 2014. *Zwyczajna przeprowadzka: Pamiętniki imigrantów z Polski do Kanady (1988-2012)* [An ordinary move: Memoirs of Polish immigrants to Canada (1988-2012)]. Toronto: Canadian-Polish Research Institute.

Kowalczyk, Anna. 2018. *Brakująca połowa dziejów: Krótka historia kobiet na ziemiach polskich* [The missing half of history: A brief history of women in the Polish lands]. Warszawa: Grupa Wydawnicza Foksal.

Makowski, William B. 1967. *History of integration of Poles in Canada*. Toronto: The Canadian Polish Congress.

Merriam-Webster. 1993. *Webster's third new international dictionary of the English language*. Cologne: Könemann Verlagsgesellschaft mbH.

Ożóg, Edward L. 1984. *Polish-Canadian immigrant literature, 1918-1940*. Ottawa, Ontario: University of Ottawa. (Master's dissertation.)

Ratyńska, Barbara. (ed.). 1971. *Pamiętniki emigrantów: Kanada, nr 1-16* [Memoirs of emigrants: Canada, 1-16]. Warszawa: Książka i Wiedza.

Romaniszyn, Krystyna. 1986. Obraz Polski a świadomość narodowa: Analiza pamiętników imigrantów polskich w Kanadzie [The worldview of Poland and national consciousness: An analysis of the memoirs of Polish immigrants in Canada]. *Studia Polonijne* 10. 295-308.

Sojka, Eugenia. 2010. Twórczość polskiej diaspory w Kanadzie a kanadyjski dyskurs literacki: Prolegomena do współczesnych badań [The works of the Polish diaspora in Canada and Canadian literary discourse: Prolegomena to contemporary research]. In Buchholtz, Mirosława & Sojka, Eugenia (eds.), *Państwo, naród, tożsamość w dyskursach kulturowych Kanady* [State, nation, identity in Canadian cultural discourses], 284-334. Kraków: Universitas.

Śliz, Anna. 2009. Gdy zobaczyłem Kanadę… Nadzieje. Oczekiwania. Rozczarowania [When I saw Canada... Hopes. Expectations. Disappointments]. *Archiwum Emigracji: Studia – Szkice – Dokumenty* 2(11). 44-54.

Toelken, Barre. 1996. *The dynamics of folklore*. Revised and expanded edition. Logan: Utah State University Press.

Wańkowicz, Melchior. 1954. *Tworzywo*. New York: Roy Publishers.

Zins, Henryk. 1975. *Historia Kanady*. Wrocław: Zaklad Narodowy im. Ossolinskich.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:3.40-50 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

It's not the glass ceiling, it's the dirt floor. Socio-discursive representations of women from San Juan (Argentina) as workers based on the Previsional Inclusion Pension

Tatiana Marisel Pizarro
Conicet – Instituto de Investigaciones Socioeconómicas – Universidad Nacional de San Juan

Abstract: This study investigates the sociodiscursive representations of women of retirement age in San Juan, Argentina, alongside the senators who enacted laws pertaining to the Previsional Inclusion Plan (PIP) and the Universal Pension for the Elderly. It focuses on diverse evaluations of unpaid domestic and reproductive labor as interpreted through the lens of the PIP by employing life stories of three women: a PIP beneficiary with incomplete pension contributions, a retiree who participated in the formal labor market, and a woman in a state of pension limbo due to evolving legislation. Furthermore, the research analyzes the parliamentary and legal discourse surrounding the recognition of housewives' retirement status under the PIP, emphasizing the implications of Laws 25.994 (2004), 26.970 (2014), and 27.260 (2016). The objective is to contribute to theoretical discussions on the biographical method and Critical Discourse Analysis (CDA), offering social scientists valuable epistemic-methodological tools that enrich research within the social sciences.

Keywords: sociodiscursive representations, women, unpaid labor, retirement policy, Critical Discourse Analysis, biographical method

1. Introduction

Historically, women have participated in various paid, albeit often precarious, employment roles such as domestic workers and nannies, which have enabled them to balance public employment with domestic responsibilities. This integration into the labor market has frequently been transient, as roles within the private sphere, particularly motherhood, have generally taken precedence. Battagliola (2004: 22) articulates that women's labor encompasses a range of meanings shaped by industrialization and the dichotomy it created between market-oriented and domestic tasks, thereby marginalizing much of women's household labor (Battagliola 2004: 4-5). Consequently, domestic labor has often been rendered invisible, while *real* work has been delineated as salaried and external.

This article explores the *Retirement for Housewives* initiative through a gendered lens, scrutinizing unpaid care work within the framework of social policy regulations and Senate deliberations. Each statute functions as a regulatory instrument that influences social behavior and identity. Acknowledging that the social significance of work varies with cultural representations, this analysis endeavors to deconstruct the perception of household responsibilities as non-work (Garley 2000: 53). This study employs Critical Discourse Analysis (CDA), a methodological and theoretical approach that explores how discourse structures reflect and reinforce relations of

power, inequality, and ideology in society (Burman & Parker 1993; Fairclough & Wodak 2000).

As Butler (2004) posits, social policies are inherently political, utilizing language to transform social entities into political agents.

2. That border between work and non-work

Recently, a growing consensus among scholars has emerged regarding the categorization of women as workers, whether they engage in labor within or outside the home. Nevertheless, there is a shared agreement that comprehending the underlying causes of women's subordination is crucial for mitigating the oppression they encounter (Benston 1969; Morton 1971; Tilly & Scott 1978; Delphy 1982; Picchio 1994; Schweitzer 2000; Rodríguez Enríquez 2011; Federici 2013 among others). Notably, Margaret Benston emphasizes that the unpaid labor performed by women constitutes a significant benefit to those who possess the means of production. She posits that housewives enable their husbands – who are wage-earning workers – to fulfill their professional roles. Consequently, Benston proposes that women act as a reserve army for the capitalist market (Benston 1973; Borderías & Carrasco 1994).

In this context, Delphy (1982) advocates for recognizing women's role within the family structure, contending that men, through patriarchal relationships, appropriate the labor performed by women – framing this work as productive labor. This gendered division of labor, linked to industrialization, positions men's work as both productive and public, while women's work is relegated to the private realm of dependency and caregiving. The *male breadwinner* model delineates family dynamics with a male provider and a female caregiver, accentuating her purported *inactive* status. This dependency further contributes to women's invisibility and their limited access to stable employment, thereby relegating them primarily to informal jobs to balance domestic responsibilities and caregiving duties. The concept of *active inactivity* associated with housewives encapsulates their social role, which fosters identity and a sense of belonging (Casas 1988).

Paid employment, often perceived as the cornerstone of public identity, facilitates social protection and future planning (Castel 2004). The redistributive structure of social security thus highlights gender inequality, compensating individuals who have invested years in unpaid caregiving roles (Sarasúa & Gálvez 2003). A more nuanced examination of caregiving as a legitimate form of labor is imperative, necessitating a re-evaluation of caregiving as a social practice that fulfills essential needs (Finch & Groves 1983).

Harrison (1975) contends that domestic labor functions as a flexible labor reserve for capitalism, allowing for fluid entry and exit from the workforce while undermining the bargaining power of male labor by providing a low-cost labor alternative – akin to an industrial reserve army.

3. Who cares for those who care?

For years, biparentalism was the manifestation of a hegemonic family, with the head of the household providing his labor force to a market willing to compensate him with money, and a housewife offering her labor to domestic and caregiving tasks, for which she receives no payment (Razavi 2010). This is key to understanding the trajectory that care has taken and who performs it to determine whether it is made visible – or not – by social policies.

During the 1980s in Latin America, the provision of social services was subject to fiscal balance, with the aim of rationalizing social spending and the management of social policies

(Gómez Urritia 2010). As Arriaga (2006: 11) expresses, the State shifted the responsibility of "finding and solving problems of illness, unemployment, physical and mental incapacity, and the death of family members" to the market, families, or women. This led to women either abandoning their responsibilities or seeking precarious jobs to prioritize family needs.

Care work has received scant social recognition, and the redistribution of this work is proportional to that value. Therefore, in line with the purposes of this article, a discursive construction of the *I, worker* of the housewife is proposed, based on Argentine social policies. To achieve this, two measures were selected, both from different governments: the Pension Inclusion Plan and the Universal Pension for the Elderly.

The debate of interest in this writing lies in how these measures were implemented, as well as the State's discourse regarding their application, effectiveness, and limitations. It is important to clarify that these reflections are made through the development of a critical process in which the concepts present in these measures are reinterpreted, considering that they were created by a patriarchal system.

4. Law 25.994: Early retirement benefit – Pension Inclusion Plan

On December 16, 2004, the project of law (CD 124/04) was unanimously approved in the Argentine Senate, thus enacting Law 25.994 (Senado y Cámara de Diputados de la Nación Argentina, 2004). With this, the creation of the Early Retirement Benefit was established, which was partially promulgated on December 29 of the same year. This law provided for an early retirement regime for workers who could prove 30 years of service and were over 60 years old for men and over 55 for women, in accordance with what was stipulated by Law 24.241 – the National Law of the Integrated Pension and Retirement System. The duration of this moratorium was set for two years, with the possibility of extension by the executive in case of justification. This law constitutes an important step in social policies, as it has a universalist approach, serving as a corrective measure for the consequences resulting from the neoliberal measures adopted during the Menem administrations. In other words, it was a coverage plan specifically designed for those who were unemployed and who had not met the age requirements stipulated by Article 19 of Law 24.241.

Law 25.994 was approved unanimously and without abstentions – both in Congress and the Senate – allowing us to infer that consensus was reached regarding the need for greater pension coverage. This is because during the 1990s, the pension coverage rate for the inactive population decreased due to: a) high unemployment levels, b) labor precariousness/informality, and c) pension reform leading to the creation of private pension funds (AFJP).

A notable fact is that this pension moratorium was one of the most successful social policies in terms of its impact on gender equity concerning access to social security, with 73% of the total benefits awarded going to female beneficiaries (ANSES 2010: 6). The paradox here is that over the years, this moratorium has commonly been referred to as the *housewives' pension*, even though it does not mention women, or the domestic and caregiving labor performed at home.

5. Law 26.970: Integrated Argentine Pension System – 2nd phase of the Pension Inclusion Plan

On July 2, 2014, Law 26.970 (O.D. No. 244/14) was enacted (Senado y Cámara de Diputados de la Nación Argentina, 2014), allowing for the regularization of pension contributions from the

period 1993-2003 through a payment plan of 60 installments, primarily aimed at self-employed workers and single-tax contributors approaching retirement age – 65 years for men and 60 years for women. This is a critical point, as it enabled self-employed workers, whether registered in the Integrated Argentine Pension System (SIPA) or not, to access the provisions of this law.

To qualify for this benefit, it was sufficient to have canceled one installment of the debt regularization regime and for the National Social Security Administration (ANSES) to determine the right to this benefit after conducting a socioeconomic and asset evaluation to ensure these provisions benefited the most vulnerable sectors of the population. This stimulated a redistributive impact, as it was a measure aimed at guaranteeing pensions and, consequently, medical coverage for individuals who could not afford it due to the high costs of pension debts. Moreover, the moratorium allowed for an increase in tax revenue from contributions, which contributed to the financial sustainability of the system (ANSES 2010).

The key point to highlight in this section is that this moratorium – similar to its predecessor – does not reference the working woman as a beneficiary but employs the generic masculine to refer to the set of beneficiaries (Articles 1, 2, 3, 4, 6, and 10). However, during the ordinary session on July 2, 2014, in which the law was discussed, there was constant reference to housewives, along with their recent visibility by the State and a recognition of the tasks performed.

6. Law 27.260: Pension for the elderly – Historical reparations law

On June 29, 2016, Law 27.260 (Senado y Cámara de Diputados de la Nación Argentina, 2016), also known as the Historical Reparations Law, was enacted. This is one of the first measures adopted in the pension sector by the neoliberal government of Mauricio Macri, and replaces the Pension Inclusion Plan implemented by the Kirchnerist administration.

The main points of divergence include the fact that this is not a retirement pension but a non-contributory pension; there is a five-year increase in the retirement age for women, it does not generate pension rights, the amount of this pension is 80% of the minimum retirement pension, and there is incompatibility with other benefits such as retirement or pension.

Like the previous regulations, this law also employs the generic masculine. This can be observed in references to *workers* (Articles 5, 10, 12, 16, and 55) and *beneficiaries* (Articles 1, 3, 4, 7, 9, 17, 18, 28, 30, and 36). The only mention of women appears in Article 22, which refers to the previously mentioned increase in retirement age.

Despite this, during the special session on June 29, 2016, where the law was discussed, there was a constant reference to the benefits brought to the adult population by the Pension Inclusion Plan, especially in relation to housewives, recognizing them as direct recipients of these benefits.

7. Analysis of selected discourses

The discursive exploration is based on three categories of analysis: subjects, work, and forms of power. These categories allow for the tracking of the discourses present throughout the textual corpus. It is important to note that in pursuit of expositional clarity, the laws and their corresponding session journals were addressed separately, despite their complementary relationship.

7.1 Discursive constructions

The six analyzed texts – three laws and their corresponding session journals from the National

Senate – refer to various subjects, from which those relevant to the study's objectives were selected for this article. In this section, one can observe how subjects – women/men and their roles as workers – are discursively constructed.

7.2 The laws

As observed, the aforementioned laws make no reference to gender as an adjective of work; representations link it to the tasks carried out by men in the labor market (the workers). This discursive construction, accomplished through omission or absence, impacts notions of identity and power: the female gender is inferred, taken for granted, and rendered irrelevant to work. In none of the analyzed legal texts are women named, suggesting that they are apparently not the recipients of these discourses. They are only considered as interlocutors and as recipients of the moratoriums in the spoken discourses during the corresponding ordinary sessions, as will be shown later.

In examining the mentioned documentary sources, gaps/omissions were observed regarding mentions of female workers, which allow for a focused discursive analysis of the written texts – Law 25.994, Law 26.970, and Law 27.260. Burman & Bunn (2011) explain that each text is developed within a historical, social, cultural, and political context, leading to a reflective practice that encourages giving meaning to what is unnamed (silences), as well as to what is present; that is, each discourse manifests absences and/or may omit others, or in this case, other groups.

Although the majority of benefits were conferred upon women, the described laws do not mention them; rather, they resort to the generic masculine to refer to both sexes, such as *the workers* and the *beneficiaries*, which is a clear example of androcentric bias in language (Eichler 1991). In other words, the generic masculine appears as an ontological category of a natural order. By opting to present subjects as a *whole*, the laws do not deconstruct either sex or gender, but rather depict an immutable, indefinable, and indecipherable being. Without underestimating the expertise of those drafting the laws, it is necessary to highlight some consequent points arising from these silences and generic language usage.

The decision to use the generic form leads to a decisive, fixed cutoff. Furthermore, by using the masculine as encompassing all subjects, there is a risk of granting certain status to men while being reductive towards women. In other words, treating gender as something neutral leads to equating the unequal, which in turn exacerbates inequality (Dema Moreno 2000). Thus, an abstract category is transformed into a material entity that dictates exclusions, as it restricts possibilities to only one: the worker. Consequently, by rendering women elliptical through absence or silence, the heteropatriarchal system is reproduced, illustrating that it is patriarchy that hierarchizes and creates genders (Amorós 1992).

The failure to name women in any of the texts indicates that policies have yet to surpass this perspective and do not propose alternative conceptual or epistemic frameworks. Moreover, this omission does not promote visibility regarding their contributions to the capitalist system or even their recognition.

It is crucial to address this analytical point, as it sheds light on the social and historical complexion of subjectivities that influences society by promoting them. In this instance, the state constructs, through omission and absence – silences – the concept of non-worker.

Thus, one can observe how language affects the ways in which citizens navigate their identities as men or women, and how those who remain unnamed are identified in opposition and contrast. The use of the generic masculine also leads to a linear approach, sidelining the distinctive and complex, thus hindering classification and even the definition of groups.

7.3 Session journals

In the specific case of the journals of the ordinary – and special – sessions for Laws 25.994, 26.970, and 27.260, sex and gender are identified as complementary and even derivative. The biological is presented as determinant of the cultural, and the individual of the social. The construction of identities (woman/man/others) determines social relationships, specifically labor relations (worker/employee).

Another point to analyze is that in the various discourses, reference was made to the imperative nature of revaluing the role of women – but, despite this, it is not detailed in the law. In this sense, it is again necessary to mention the silences surrounding the discursive construction of female workers.

It is also interesting to observe and characterize the discursive construction of men, how they present women, and the derived consequences. In this regard, various senators made efforts to demonstrate a commitment, defense, and promotion of the recognition of housewives and their labor by advocating for changes in certain sociocultural patterns. This can also be seen in that those who referred to the moratorium did so by using the colloquial term for it: *housewife pensions*.

> Now, in this unregistered work, women have suffered the most, which was clearly demonstrated by the events of the previous moratorium, where over seventy percent of those who benefited from that moratorium were, in fact, women. Therefore, I am convinced that this should be one of the banners that we, as women, defend; it should be one of the major issues that should have been addressed even in the Women's Bank. I read on the backs of some of the women up there that being a housewife is work. And indeed, it is. I believe that there is no other job that deserves greater recognition from society than the work of a housewife, not only for all the tasks she performs in the home but also for her contributions to children's values, education, and society as a whole. We are so convinced of this that in 1993, during a time of privatization and when other winds were blowing, Law 24.476 was enacted, originating from project file 4.503/93. This was the first project presented for the retirement of housewives and a moratorium to grant pensions to housewives. We are certainly convinced of this. (Senator Silvia Elías de Pérez, Diario de sesiones de la Cámara de Senadores de la Nación, July 2, 2014)

> I cannot forget about the pensions for housewives because it brought so much satisfaction to my province. There were so many women who never imagined in their lives they would be able to receive a pension for their work as housewives. Additionally, we had non-contributory pensions for those over 70 years old. I believe that was indeed a historical reparation. (Senator Liliana Fellner, Diario de sesiones de la Cámara de Senadores de la Nación, June 29, 2016)

As can be observed, the senators emphasized in their presentations the invisible tasks performed by women within the household, their constant socioeconomic vulnerability, and the recognition as *workers* granted through this moratorium. This is in stark contrast to the first measure – Law 25.994 – where neither the text of the regulation nor the session made any mention of women.

> I specifically refer to women housewives. Because they normally appear in all statistics, when a survey is conducted in a household, it shows that Mr. So-and-so

> works, is an accountant, and when asked about the woman, if she is a housewife, it is stated that she "does not work." Yet that woman performs a task that has no fixed hours, takes care of the children, washes everything that needs to be washed after meals, sets the table, cleans the house, and she never gets paid... after 2003, the same criteria were applied because women who could not contribute were workers who were not compensated; they did not appear in the gross product and engaged in silent work without which it would be impossible to imagine society functioning normally. So, it is a profound act of justice that this pension without contributions for women appears and that these women are no longer from one or two provinces, but rather millions of women who now have access to an act of justice. (Senator Salvador Cabral, Diario de sesiones de la Cámara de Senadores de la Nación, July 2, 2014)

In this regard, the constant presence of discursive constructions surrounding the housewife and references to her labor may signify some progress in the socioeconomic recognition of care and domestic tasks as work, yet not enough to be explicitly expressed in the texts of the regulations.

Moreover, it is evident that there are those who try to conform to the use of these discursive constructions; however, what they ultimately do is disseminate the false ideal that the tasks of the housewife have monetary value, which is only acknowledged when performed in the public sphere.

> We all agree on recognizing the work they may have had, some formally, others informally; some who lost their jobs or those who worked partly formally and partly informally. Also, beyond housewives, many people have worked making pastries, selling from their homes, or as street vendors, even those who performed entertainment at stoplights to earn some money... The housewife washing or ironing for someone else or the street vendor, they all worked. (Senator Roberto Basualdo, Journal of Sessions of the Senate of the Nation, July 2, 2014)

In line with this argument, one can also observe a discursive construction that erects the perception of a woman dependent on the state, a passive recipient of aid that is not truly transformative, as it does not fully emancipate her nor emphasize her autonomy:

> In this case, we are speaking of a present state that understands it must take responsibility, that it must provide support, and that in some way it must facilitate for that sector a right as important as retirement. Not to mention the amount of the pension. I still remember my mother going to collect her pension and returning with coins, separating them to pay each of her bills, and that was all she had. (Senator María Higonet, Journal of Sessions of the Senate of the Nation, July 2, 2014)

From the implementation of the first moratorium to the current Pension for the Elderly, the speeches presented display greater awareness and recognition of the labor and role of the housewife within the home; a change in the representations of the definition of work can be perceived. Although this opens a new field of intervention that goes beyond the private sphere of the home, the issue of unpaid care and domestic work is still not specifically addressed or regulated.

It is also true that the discursive constructions extracted from the aforementioned sessions criticize the housewife as an antonym of citizenship, as the former pertains to the private domain,

while the latter pertains to the public. While the speeches attempt to establish a connection between the public and the private, it is inferred that the domestic sphere – private – hinders the exercise of citizenship – public.

Finally, it is imperative to emphasize that in the sessions, women were mentioned in their role as housewives, yet none of the senators alluded to the realities of unemployment, underemployment, and precarious labor that they face in their economically active lives, which also entails their inability to complete their pension contributions. Thus, by discursively associating the work of women directly with caregiving and reproductive tasks, one inevitably falls into essentialist views of a patriarchal ideal of femininity. In this manner, these constructions imply that being a housewife is a constitutive part of a woman's identity rather than an external aspect of her, a consequence of the realities, decisions, or impositions accepted upon assuming this role. Within these discourses, there is a singular representation of womanhood that disregards any reference to autonomy outside the familial domain. That is to say, there is no allusion to someone emancipated and free, but rather to a submissive housewife enslaved by household chores.

The evolution of the discursive constructions regarding housewives, their work, and their rights has followed a trajectory resembling that of a boomerang. It started from nothing, from a law that did not consider them at any moment – neither in its written text nor in the session in which it was addressed. It progressed toward a positive externality that acknowledged them as citizens, workers, and deserving of rights as such.

Perhaps this regression is partly due to the fact that these discourses are still constructed within the framework of the dichotomy *private spaces versus public spaces.* The discursive constructions are thus erected with the cultural ideal of the housewife who provides her labor within the private sphere of the home, accompanied by a notable absence in the mentioned measures; her absence has become normalized. It can be observed, then, that omissions, silences, and words are never neutral or innocent.

8. Final reflections

In the analyzed texts – laws and session transcripts – the unsaid, expressed through silence and omissions, establishes a particular way of viewing and understanding our surroundings. In this context, the silences highlight the lack of interrelation between the public and private spheres, particularly concerning women, their work as housewives, and their recognition as workers deserving rights.

In light of the above, it can be observed that both the Pension Inclusion Plan, its second phase, and the Universal Pension for the Elderly have been conceived without a gender perspective, addressing a generic collective of workers while failing to consider the specificities and unique circumstances of women – or any other group. It should not be forgotten that it is gender, which, in a patriarchal society, becomes the social norm; it is these sex-gender indicators that hierarchically structure society in a patriarchal manner (Amorós 1992).

Rather than expanding the scope of application with detailed specifications of the different groups targeted by these measures, they obfuscate interpretation by rendering women invisible – thus leaving them unprotected. That said, while such issues are manifested in the texts of the laws, the ordinary sessions emphasize the discursive constructions related to housewives and the significance of their unpaid labor, thus underscoring the necessity for recognition through pension benefits. The discourses presented in the sessions exhibited changes: housewives are not merely engaged in chores out of love; rather, it is unremunerated labor. The debt that the state owes to

them was acknowledged.

This article aimed to consider how these three social policies are intertwined with three dimensions, despite not explicitly targeting them: the political aspect – as the sexual division of labor underlies (Picchio 1992); the structural dimension – due to the existence of the figure of the dominant/dominated; and the interpersonal aspect – through which collective social experiences also reproduce subjectivities.

Through this critical discourse analysis, an attempt was made to demonstrate how the discursive constructions surrounding these moratoria allowed for the creation of interpretative frameworks and subtly intervened in the valuation of housewives' tasks as work. With its implementation, the Pension Inclusion Plan served as an interpretative guide to reality, endorsing a new configuration of representations and discursive constructions, along with new positionalities.

It is also imperative to emphasize that one of the main consequences of the silence/omission of women in the texts of the laws is the continued confinement of their roles to the private sphere, rendering their work and silent contributions invisible to the state.

Unintentionally, in the sessions concerning these regulations – after nearly 15 years – it became evident that some structures of meaning are being altered, which in the future may allow for greater social action and an impact on those who benefit from these measures. Nevertheless, words can easily vanish, and it is essential that laws are drafted with a gender perspective and that this perspective is enshrined in writing.

References

Agencia Nacional para la Seguridad Social –Anses– 2010. *Análisis de la Cobertura Previsional del SIPA: Protección, Inclusión e Igualdad*. https://www.anses.gob.ar/sites/default/files/2022-05/DT_Análisis%20de%20la%20cobertura.pdf

Amorós, Celia. 1992. Notas para una teoría nominalista del patriarcado. *Asparkía: Investigación feminista* 1. 41-58. https://www.e-revistes.uji.es/index.php/asparkia/article/view/412

Arriaga, Isabel. 2006. *Género y políticas públicas: nuevas perspectivas para el análisis de la equidad*. México: UNIFEM / El Colegio de México.

Battagliola, Fabiana. 2004. *Histoire du travail de femmes*. Paris: La Découverte.

Benston, Margaret. 1969. The Political Economy of Women's Liberation. *Monthly Review* 21(4). 13-27.

Benston, Margaret. 1973. Para una economía política de la liberación femenina. In Capmany, Maria Aurelia & Balletbó, Ana & Falcón, Lidia et al. (eds.), *La liberación de la mujer: Año Cero*, 33-46. Buenos Aires: Granica.

Borderías, Cristina & Carrasco, Cristina. 1994. Las mujeres y el trabajo: aproximaciones históricas, sociológicas y económicas. In Borderías, Cristina & Carrasco, Cristina & Alemany, Carmen (eds.), *Las mujeres y el trabajo,* 15-110. *Rupturas conceptuales.* Barcelona: Icaria.

Burman, Erica & Bunn, Graham. 2011. *Discourse Analysis.* Reino Unido: Manchester Metropolitan University.

Burman, Erica & Parker, Ian. 1993. *Discourse analytic research: Repertoires and readings of texts in action*. London: Routledge.

Butler, Judith. 2004. *Lenguaje, poder e identidad.* Madrid: Síntesis.

Casas, Lidia. 1988. *La mujer y el trabajo doméstico: un estudio socio-jurídico sobre el trabajo doméstico no remunerado*. Santiago de Chile: FLACSO.

Castel, Robert. 2004. *La inseguridad social ¿Qué es estar protegido*? Buenos Aires: Manantial.

Delphy, Christine. 1982. *El enemigo principal. En Por un feminismo materialista. El enemigo principal y otros textos*. Barcelona: La Sal.

Dema Moreno, Silvia. 2000. *A la igualdad por la desigualdad: la acción positiva como estrategia para combatir la discriminación de las mujeres.* Buenos Aires: KRK.

Diario de sesiones de la Cámara de Senadores de la Nación. 37° Reunión – 30° Sesión Ordinaria, 16 de diciembre de 2004. http://www1.hcdn.gov.ar/sesionesxml/item.asp?per=122&r=36&n=29

Diario de sesiones de la Cámara de Senadores de la Nación. 10° Reunión – 6° Sesión Ordinaria, 2 de julio de 2014. http://www.senado.gov.ar/parlamentario/sesiones/busqueda

Diario de sesiones de la Cámara de Senadores de la Nación. 10° Reunión – 7° Sesión Especial, 29 de junio de 2016. http://www.senado.gov.ar/parlamentario/sesiones/busqueda

Eichler, Marcia. 1991. *Nonsexist research methods: A practical guide.* London: Routledge.

Fairclough, Norman & Wodak, Ruth. 2000. Análisis crítico del discurso. In van Dijk, Teun A. (ed.), *El discurso como interacción social*, 367-404. Buenos Aires: Gedisa.

Federici, Silvia. 2013. *La revolución en punto cero. Trabajo doméstico, reproducción y luchas feministas.*Traficante de Sueños. (Revolución en punto cero. Trabajo doméstico, reproducción y luchas feministas)

Finch, Janet & Groves, Diane. 1983. *A labour of love: Women, work and caring.* London: Routledge.

Garley, Deborah. 2000. Perspectivas históricas. In Maruani, Margaret (ed.), *Las nuevas fronteras de la desigualdad. Hombres y mujeres en el mercado de trabajo,* 35-58. Barcelona: Icaria.

Gómez Urrutia, Alicia. 2010. *Estado y política social en América Latina: una mirada crítica desde la economía feminista*. Buenos Aires: CLACSO.

Picchio, Antonella. 1994. El trabajo de reproducción, tema central en el análisis del mercado de trabajo. In Borderías, Cristina & Carrasco, Cristina & Alemany, Celia (eds.)*, Las mujeres y el trabajo: rupturas conceptuales*, 451-490. Barcelona: Icaria.

Razavi, Shahra. 2010. Unequal worlds of gender and care: The journey of feminist ideas into the policy realm. *Ponencia en el Seminario Internacional "Políticas de cuidado, género y bienestar," 14 de octubre.* Buenos Aires: UNICEF.

Rodríguez Enríquez, Corina. 2011. *Programas de transferencias condicionadas de ingreso e igualdad de género. ¿Por dónde anda América Latina?* Serie Mujer y Desarrollo. Buenos Aires: CEPAL.

Sarasúa, Carmen & Gálvez, Lina. 2003. *¿Privilegios o eficiencia? Mujeres y hombres en los mercados de trabajo.* España: Universidad de Alicante.

Senado y Cámara de Diputados de la Nación Argentina (29 de diciembre, 2004). Ley 25.994, Créase una prestación previsional anticipada. Requisitos que deberán cumplir las personas que tendrán derecho al mencionado beneficio previsional. Presidencia de la Nación. SEGURIDAD SOCIAL Ley 25

Senado y Cámara de Diputados de la Nación Argentina (9 de septiembre, 2014). Ley 26.970, Régimen de Regularización – Sistema Integrado Previsional Argentino. Presidencia de la Nación. InfoLEG - Ministerio de Economía y Finanzas Públicas - Argentina

Senado y Cámara de Diputados de la Nación Argentina (26 de mayo, 2016). Ley 27.260, Créase el Programa Nacional de Reparación Histórica para Jubilados y Pensionados. Presidencia de la Nación. http://servicios.infoleg.gob.ar/infolegInternet/anexos/260000-264999/263691/norma.htm

Tilly, Louise & Scott, Joan. 1978. *Women, Work, and Family*. New York: Holt, Rinehart and Winston.

This article was first published at https://www.lacussquare.org

LACUS Forum **50:3.51-60 (2025)**

LACUS Forum is published on-line at *lacus.weebly.com*, where an electronic copy is provided, free of charge and with no implied warranty. Its contents are available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0; see *http://creativecommons.org/licenses/by-nc/3.0/*. Copyright © 2025 the Linguistic Association of Canada and the United States.

Two birds, one stone – idioms across languages and the language brokering effect

Joyce Zeng
McMaster University

Abstract: The article examines language brokering in highly proficient Chinese-English bilinguals. Language brokering involves informal translation to facilitate communication between two parties in two languages. Previous research shows that language brokers tend to be more agile across languages. The study used a counterpart idiom task to test language brokers' and non-language brokers' comprehension accuracy in both English and Chinese language translation directions, assessing idiom type with respect to decomposability considering decomposable (D) idioms such as *kill two birds with one stone* and its Chinese equivalent 一石二鸟 (two achievements with one action) and non-decomposable (ND) idioms such as *not all roses* and its Chinese equivalent 不尽如人意 (not entirely as one wishes). Results showed language brokers performed similarly across D and ND, while non-language brokers scored higher on D than on ND. Overall, language brokers had higher accuracy, performing more consistently across languages and idiom types.

Keywords: language brokering, bilingualism, counterpart idioms, idiom types, communication

1. Introduction

CANADA'S LINGUISTIC LANDSCAPE is constantly evolving as immigrants from diverse cultural backgrounds bring their native languages and lived experiences into a multilingual society. However, many immigrants are not fully equipped to navigate the cultural and linguistic demands in their new environment. This is particularly relevant to a new immigrant's understanding of the host nation's figurative language use (e.g. idioms), which often presents unique challenges due to their non-literal meanings.

In most cases, the duty of bridging this language gap falls to younger generations who are more linguistically agile in their immigrant community (McQuillan & Tse 1995; Weisskirch 2010). They have acquired the host country's language through formal education and daily natural exposure, enabling them to translate seamlessly from one language to another and vice versa. With this skill, they often act as language brokers (LBs), who are individuals bridging communication between two parties in two different languages (Katz 2014). Language brokering activities are typically spontaneous in everyday settings such as medical appointments, legal consultations, school communiques, or even restaurant menu explanations (Buriel et al. 1998; Weisskirch 2006, 2010; Vaid et al. 2008; Katz 2014; López & Vaid 2016, 2018; López et al. 2019; López 2020). Thus, LBs have the chance to experience repeated practice and can develop an enhanced cognitive flexibility and

metalinguistic awareness.

Most existing studies on language brokering have focused on second-generation children of immigrant families (Tse 1996; López et al. 2019; Weisskirch 2010). However, the present study examines a demographic that has received less attention. The study focuses on highly proficient young adult bilinguals who have recently immigrated to Canada to pursue post-secondary education from China. These young adults often act as language brokers for their older immigrant relatives who have joined them in Canada to support their academic goals but may not have the language skills necessary to communicate independently (Bastien & Tuey 2023).

2. Literature review and theoretical framework

This study aims to build on existing work in the field of LBs and idioms to examine the cognitive and linguistic outcomes of informal translation practices (Tse 1996; García et al. 2014; Rainey et al. 2015; López et al. 2017; López & Vaid 2018; López 2020).

Language brokering being a form of informal translation, does not follow the same rules, standards, or procedures as formal translation does (Tzou et al. 2017). The emphasis is not on literal accuracy but more on facilitating communication across linguistic and cultural boundaries. Moreover, LBs often rely on strategies such as paraphrasing, code-switching, and context-based inference to interpret meaning when communicating in real-time. The linguistic skills acquired may also include increased cognitive flexibility and a greater sensitivity to figurative language use (López & Vaid 2018). The positive feedback from acting as LBs and its subsequent outcomes (e.g. higher grades at school, etc.), further results in increased self-confidence and self-esteem (Buriel et al. 1998; Weisskirch 2006; López et al. 2019).

Nevertheless, figurative language use in any language can present a significant challenge for bilinguals due to its non-literal and culturally embedded meanings. Idioms are fixed expressions whose meanings often cannot be derived literally from their individual component words (Gibbs et al. 1989). In addition, the notion of idiom decomposability refers to the degree to which the literal meanings of an idiom's components contribute to understanding its overall figurative meaning (Gibbs et al. 1989). For example, a decomposable idiom (e.g. *kill two birds with one stone*) offers semantic cues that aid in meaning inference, while a non-decomposable idiom (e.g. *kick the bucket*) is more opaque and requires holistic retrieval from memory without support from its component words (Gibbs et al. 1989).

According to the Decomposability Hypothesis proposed by Titone & Connine (1999), idioms with higher semantic transparency are processed more efficiently, particularly by second language (L2) speakers who may often rely on literal analysis making decomposable idioms easier to access and interpret. On the other hand, non-decomposable idioms impose a higher processing cost due to the absence of word-level semantic cues, which need greater familiarity and direct lexical access (Gibbs et al. 1989; Titone & Connine 1999). Regardless of an idiom's semantic transparency, idioms, as figurative language, are essential to both linguistic competence and cultural influence. In this respect, idioms are frequently used to assess language knowledge (Vaid et al. 2015; López & Vaid 2018; Zhu & Minda 2021).

López & Vaid (2018) found that LBs processed English idioms faster and more accurately than non-language brokers (NLBs) when identifying their Spanish equivalents, suggesting that language brokering experience may facilitate figurative language comprehension. Nevertheless, their study focused solely on unidirectional L2 to L1 interpretation (e.g. English to Spanish) and did not examine idioms across both languages.

Conversely, language brokering in real-world settings often requires bidirectional interpretation across both languages and idiom types, including both semantically transparent and opaque idioms. Furthermore, less is known about how decomposability interacts with LB status in bilingual idiom processing across language translation directions. Do LBs process both decomposable and non-decomposable idioms more uniformly due to their frequent exposure to such forms in both languages? Are NLBs more likely to struggle with opaque idioms as predicted by the Decomposability Hypothesis? To address these questions, the present study introduces a novel bidirectional idiom equivalence task using matched English and Chinese idioms to investigate how idiom decomposability, language translation direction (English to Chinese vs. Chinese to English), and broker status interact in bilinguals' figurative language comprehension.

2.1 Research questions and hypothesis

The research questions are as follows. (1) Does idiom type affect comprehension differently for LBs and NLBs? (2) Does the interaction between idiom type and translation direction vary by broker status? Based on the Decomposability Hypothesis (Titone & Connine 1999), it is hypothesized that decomposable idioms may show higher comprehension scores than non-decomposable idioms, expressly for NLBs, who may rely heavily on semantic transparency (Gibbs et al. 1989; Titone & Connine 1999). However, LBs may perform more uniformly across idiom types due to their language brokering experience (Lopez & Vaid 2018).

Moreover, it's further hypothesized that broker status may be affected by the translation direction. LBs are expected to show stable performances in both language translation directions, reflecting their bidirectional language brokering experience. In contrast, NLBs are expected to show different performances depending on language translation direction consistent with prior findings on second language processing where the studies had indicated translation from L2 to L1 was generally more accurate than L1 to L2 (Kroll & Stewart 1994; Jiang 2000).

3. Methodology

3.1 Participants

Participants were Chinese-English bilingual young adults (N = 38) whose first language was Chinese, and where the age ranged from 18 to 40. Broker status was classified based on their self-reported language brokering experience (LBE) either as LBs or NLBs. Previous research supports the validity of self-reporting as a reliable representation when corroborated with task performance (Flege et al. 2002; Dunn & Fox Tree 2009).

Further eligibility criteria included having current or previous enrollment in a Canadian post-secondary institution. English proficiency was demonstrated by a minimum IELTS (International English Language Testing System) score of 7, while for Chinese proficiency, participants were required to have completed the Gaokao (Nationwide Unified Examination for Admission to General Universities and Colleges), which is China's national college entrance exam. By having both measures, participants were deemed bilingual at a high proficiency in both languages.

The final sample size included 19 LBs (8 males, 11 females; M = 26 years old; SD = 6.03), and 19 NLBs (9 males, 10 females; M = 22 years old; SD = 2.72). Participants were recruited through a combination of methods, including social media platforms (e.g. WeChat), printed posters with the QR code, and the university's SONA research participation system. The posters were displayed across McMaster University and in the SONA system. Participation was either voluntary or

incentivized via course credit if they were a current McMaster student.

3.2 Design

The experiment employed a mixed factorial design and was developed using online platform LimeSurvey.[1] Ethical approval was obtained from the McMaster University Research Ethics Board (MREB). The study included one between-subjects factor [broker status: language brokers (LBs) and non-language brokers (NLBs)] and two within-subjects factors [idiom type: decomposable (D) and non-decomposable (ND); language translation direction: English to Chinese (EC) and Chinese to English (CE)]. The mean comprehension score was the dependent variable measured through the multiple-choice idiom comprehension task. The design enabled the investigation of both main effects and interaction effects of idiom type, language translation direction, and broker status on idiom comprehension performance.

3.3 Materials

A multiple-choice idiom comprehension task was conducted, containing 64 counterpart idiom trials. These included 32 idioms translated from Chinese to English and 32 from English to Chinese. Within each translation direction, half of the idioms were classified as decomposable and the other half as non-decomposable (16 items per idiom type).

The final set of stimuli was selected based on two pretests assessing idiom frequency and cross-linguistic equivalency. First, 150 English idioms were drawn from the *Cambridge International Dictionary of Idioms*[2] and rated by 10 native English-speaking graduate students from McMaster University. Participants rated the frequency of daily idiom use on a 5-point Likert scale (1 = not at all, 5 = extremely). Idioms with average ratings over 60% were classified as high-frequency and shortlisted for inclusion into the task.

Non-decomposable (Chinese to English)	Decomposable (English to Chinese)
小题大做	Blood is thicker than water
Please choose only one of the following:	Please choose only one of the following:
° Have fun and games	°血脉喷张
° Be as fresh as a daisy	°血流漂橹
• Storm in a teacup	°血流成河
° Be full of beans	•血浓于水

Figure 1: Task example from the experiment (correct answer shown by •)

Additionally, 150 Chinese idioms were selected from *500 Common Chinese Idioms*: *An Annotated Frequency Dictionary* by Jiao et al. (2011). To ensure translation equivalency between English and Chinese, candidate pairs were cross-validated using the *Cambridge Dictionary of English-Chinese Translation,* a project conducted and validated by Zhu Zhu and Helen Freeman in 2022 at Cambridge University. This online platform provided idioms and their equivalents in the other language. Also, a group of highly proficient Chinese-English bilingual graduate students at McMaster University

[1] https://www.limesurvey.org

[2] https://archive.org/details/Cambridge-international-dictionary-of-idioms

independently rated each pair by responding to the question: "Are these idioms equivalent in meaning across both languages?" For example, the English idiom *practice makes perfect* was matched with the Chinese 熟能生巧 (skill comes with practice). Only those idiom pairs that received unanimous "yes" responses were retained. Based on the dual criteria of high frequency and confirmed cross-linguistic equivalency, a final set of 64 idiom trials was selected for the experiment as per Figure 1.

3.4 Procedure

The task was conducted in a non-laboratory setting using the online program LimeSurvey. Participants accessed the experiment individually by either clicking the weblink or scanning the QR code. Upon entrance to the task, the screen first displayed a digital informed consent form outlining the study's purpose, task instructions, potential risks, confidentiality measures, and the voluntary nature of participation. Participants were asked to read the form carefully and provide consent before commencing the task by clicking on the "agree" button.

Participants were instructed to complete the task at a time of their choosing. During the task, each trial displayed a single idiomatic phrase and participants selected the most equivalent interpretation from four multiple-choice options. They were instructed to respond as accurately as possible. Only response accuracy was recorded.

Lastly, after participants completed the task, they were progressed to the second section which included a brief demographic questionnaire where they self-rated their own proficiency in Chinese and English, the length of language brokering experience (e.g. if they were LBs), and basic demographic information (e.g. age, gender, education, birth country). The entire experiment time was to take approximately 50 minutes.

3.5 Data analysis

Data was analyzed utilizing IBM SPSS (Version 30). Inferential and descriptive statistics were generated to examine the distribution of comprehension scores across all conditions. Outliers were removed through visual inspection using boxplots which amounted to three participants. The reasoning was that extremely fast responses (e.g. under 20 minutes) were considered potentially invalid as it may not reflect true idiom recognition and therefore the three were removed. Thus, the final reported results reflect analyses conducted on the final clean dataset.

4. Results

4.1 Comprehension score by idiom type, language translation direction, and broker status

A three-way mixed ANOVA was conducted to investigate the effects of Idiom Type (D and ND) x 2 Language Translation Direction (CE and EC) x 2 Broker Status (LB and NLB) on idiom comprehension accuracy scores. The results showed a significant three-way interaction effect among idiom type, language translation direction, and broker status, ($F(1, 36) = 4.305$, $p = .045$, partial $\eta^2 = .107$, indicating that the effect of idiom type on comprehension scores varied depending on both the language translation direction and participant's brokering status. As seen below in Figure 2, both LBs and NLBs scored higher on D than ND with LBs performing more uniformly across language translation directions and idiom types, while NLBs show a larger drop in scores, especially for ND translated from EC.

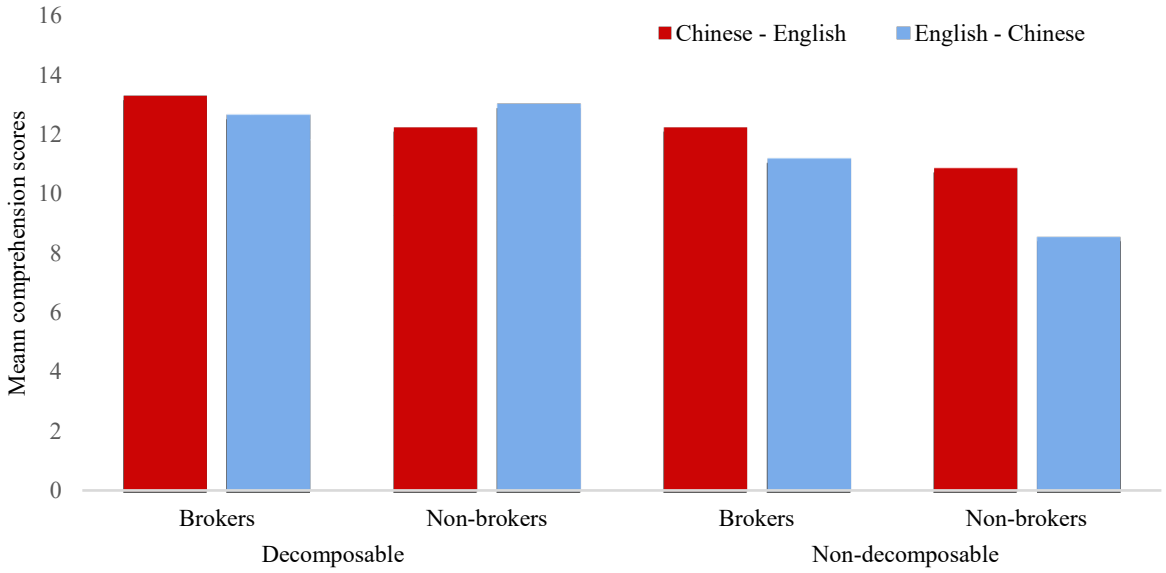


Figure 2: Comprehension scores by idiom type, language translation direction, and broker status

4.2 Idiom type accuracy depends on language translation direction for both groups

To further explore the three-way interaction, a simple two-way interaction between idiom type and language translation direction on comprehension scores was investigated separately for LBs and NLBs. For NLBs, the interaction was significant, $F(1,18) = 10.510$, $p = .005$, partial $\eta^2 = .369$, indicating that the effect of idiom type on comprehension scores varied depending on the language translation direction, while for LBs, the interaction was not significant, $F(1, 18) = .234$, $p = .634$, partial $\eta^2 = .013$, suggesting that comprehension scores for LBs was not affected by either idiom type or language translation direction.

4.3 Main effect of idiom type on non-brokers (NLBs)

For NLBs, a significant main effect of idiom type was observed in both language translation directions. In the CE language translation direction, idiom type played a role on comprehension scores, $F(1, 18) = 4.561$, $p = .047$. This effect was even more pronounced in the EC language translation direction, $F(1, 18) = 49.336$, $p < .001$.

Follow-up comparisons showed that in the CE language translation direction, D resulted in higher comprehension scores ($M = 12.21$, $SE = 0.53$) than ND ($M = 10.84$, $SE = 0.47$), with a mean difference of 1.37, 95% CI [0.02, 2.72], $p = .047$. Similarly, in the EC language translation direction, D ($M = 13.00$, $SE = 0.34$) scored better than ND ($M = 8.85$, $SE = 0.56$), with a larger mean difference of 4.45, 95% CI [3.32, 5.81], $p < .001$. The graph in Figure 3 below shows that NLBs scored highest on D in both language translation directions, while their performance declined markedly for ND compared to D in the EC language translation direction.

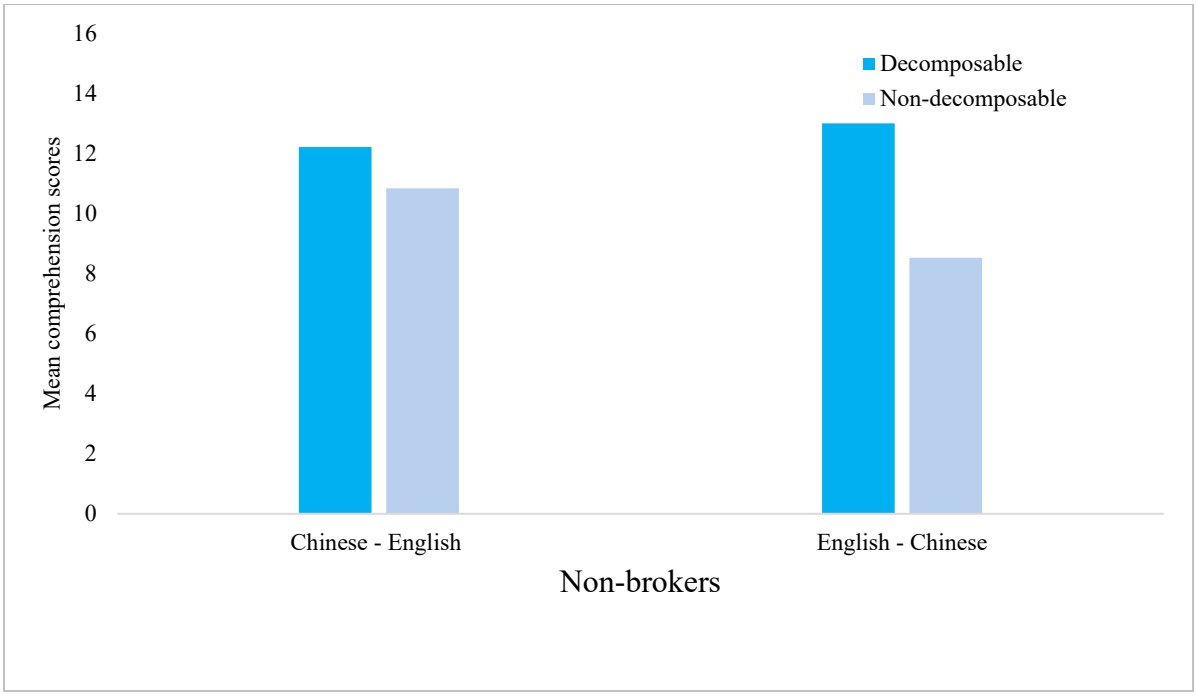


Figure 3: Comprehension scores of idiom type at language direction by non-brokers

5. Discussion

The present study investigated the effects of idiom type, language translation direction, and broker status on bilingual idiom comprehension. The two research questions were: (1) Does idiom type affect comprehension differently for LBs and NLBs? (2) Does the interaction between idiom type and translation direction vary by broker status?

Two main hypotheses were proposed whereby D would yield higher comprehension scores than ND, particularly among NLBs who may rely more on semantic transparency. Secondly, LBs were predicted to perform uniformly across both idiom type and language translation directions due to their experience in spontaneous language brokering. Therefore, the findings supported both hypotheses. Following the first hypothesis, D was scored more accurately than ND, especially among NLBs. The findings were in line with the Decomposability Hypothesis proposed by Titone & Connine (1999), thus, supporting that semantically transparent idioms facilitate comprehension by providing inferred meaning from component words. Moreover, NLBs with limited idiom exposure and less real-world translation practice (e.g. lack of brokering experience) appear to depend more heavily on semantic cues when interpreting figurative expressions.

For the second hypothesis, LBs demonstrated uniform performance across idiom types and language translation directions. This could be interpreted with their broader exposure to figurative language and frequent cross-linguistic mediation by engaging in brokering activities. Furthermore, these findings are in line with previous research which indicates that LBs develop advanced metalinguistic awareness through constant exposure in both languages, figurative language recognition, and varied settings (Vaid et al. 2015; Lopez & Vaid 2018). The absence of an interaction effect in LBs suggests that their comprehension scores were all equally high and not influenced by either idiom type or language translation direction. This consistency may be attributed to their extensive experience navigating the meanings of idiomatic expressions across languages, likely reducing their reliance on word surface semantic cues.

In contrast, NLBs showed a significant interaction effect between idiom type and language translation direction which also supports the second hypothesis, where D was scored higher than ND in both language translation directions with the strongest effects observed in the EC language translation direction. Interestingly, the strong performance by NLBs in the EC language translation direction may also suggest that NLBs rely more on semantic transparency when translating into their dominant L1 assisting them in translation efforts. This may indicate an L1 mediated decoding strategy whereby they used L1 conceptual representations to interpret L2 input in line with Jiang (2000) and Kroll & Stewart (1994), and could be interpreted as to why D in the EC language translation direction was the most accurate for NLBs.

Furthermore, NLBs seemed to be challenged the most with ND in both language translation directions, especially in CE. This likely can be explained by an unfamiliarity with English ND idioms, where the absence of transparency and not knowing its equivalent phrase in Chinese makes it visually and semantically more challenging. Despite NLBs having extended exposure in both Chinese and English with most reporting over 15 years of English learning, many NLBs still showed difficulty accessing opaque idioms without contextual or transparent structure cues. The results underscore the importance of both semantic transparency and direction asymmetry in NLB's idiom processing, where their stronger performance on D suggests a reliance on word surface semantic cues decoding rather than stored holistic representations.

5.1 Brokering experience and cognitive flexibility

Showing consistency across all conditions, LBs presented an ability to process all idiom types and language translation directions more easily, presenting greater linguistic flexibility supporting prior findings that language brokering experience enhances figurative language recognition among other positive outcomes (Tse 1996; López & Vaid 2018; López et al. 2019). This increased language exposure and practical engagement with both languages, enables LBs to have this cognitive flexibility. Thus, LBs may draw on both holistic and analytical strategies in idiom processing shaped by repeated exposure to figurative language during their language brokering activities. For example, when they encounter ND expressions through frequent exposure in real-world brokering contexts, it may be easier for them to retain and recall them in their mental lexicon. Language brokering experience through the high performance of LBs in this study has shown to be an integral factor in bilingual language advancement.

5.2 Limitations and future research recommendation

Since the task was administered online, this may have reduced experimental control compared to a supervised university lab setting. Consequently, it may be considered as a limitation. However, this limitation was mitigated through item randomization and time-stamping procedures designed to ensure the task was impervious to possible collaboration and timely completion.

Secondly, the absence of contextual cues may also present a limitation. Participants knew the test involved idioms, but no additional context was provided. Prior studies such as Siyanova-Chanturia et al. (2011) demonstrated that contextual support can improve comprehension of idioms, particularly when faced with less common or opaque idioms. Including context, such as embedding non-decomposable idioms (e.g. *kick the bucket*) within sentences or paragraphs (e.g. *he kicked the bucket in his sleep*) could have facilitated the challenges of participants when faced with ND expressions in both languages. This could be a topic of future study direction.

Another future research direction could compare language brokers and formally trained professional translators, applying real-time measures such as eye-tracking to investigate the nuanced processing dynamics underlying figurative language comprehension in both groups.

6. Conclusion

The study contributes to the literature by presenting that the processing of idioms for bilinguals is not only affected by structural properties but also real-world language exposure and translation experience, adding to language brokers' dynamism and adaptability.

Offering valuable insights into the cognitive mechanisms underlying idiom comprehension, the study further underscores the distinct role of language brokering experience in shaping cross-cultural and interpersonal communication.

As Canada's multicultural landscape continues to grow, situations requiring language brokering are becoming increasingly common. Language brokering not only reflects the fundamental human drive to communicate but also embodies the resilience and adaptability of individuals striving to connect across linguistic and cultural boundaries.

References

Bastien, Nicholas & Tuey, Catherine. 2023. *A portrait of non-permanent residents in Canada*. Government of Canada, Statistics Canada. Available online at https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2023039-eng.htm.

Buriel, Raymond & Perez, William & De Ment, Terri L. & Chavez, David V. & Moran, Virginia R. 1998. The relationship of language brokering to academic performance, biculturalism, and self-efficacy among Latino adolescents. *Hispanic Journal of Behavioral Sciences*, 20. 283-297.

Dunn, Alexandra L. & Fox Tree, Jean E. 2009. A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition* 12(3). 273-289.

Flege, James Emil & Mackay, Ian R. A. & Piske, Thorsten. 2002. Assessing bilingual dominance. *Applied Psycholinguistics* 23. 567-598.

García, Adolfo & Ibañez, Agustín &, Huepe, David & Houck, Alexander L. & Michon, Maëva & Lezama, Carlos G. & Chadha, Sumeer & Rivera-Rei, Álvaro. 2014. Word reading and translation in bilinguals: The impact of formal and informal translation expertise. *Frontiers in Psychology* 5. 1-14.

Gibbs, Raymond W. & Nayak, Nandini P. & Cutting, Cooper. 1989. How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language* 28. 576-593.

Jiang, Wen Ying. 2000. The relationship between culture and language. *ELT Journal* 54. 328-334.

Jiao, Liwei & Kubler, Cornelius C. & Zhang, Weiguo. 2011. *500 Common Chinese Idioms: An Annotated Frequency Dictionary*. New York: Routledge.

Katz, Vikki. 2014. Children as brokers of their immigrant families' health-care connections. *Social Problems* 61. 194-215.

Kroll, Judith F., & Stewart, Edward. 1994. Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language* 33(2). 149-174.

López, Belem G. 2020. Incorporating language brokering experiences into bilingualism research: An examination of informal translation practices. *Language and Linguistics Compass* 14(1). 1-19.

López, Belem G. & Lezama, Esteffania & Heredia Jr., Dagoberto. 2019. Language brokering experience affects feelings toward bilingualism, language knowledge, use, and practices: A qualitative approach. *Hispanic Journal of Behavioral Sciences* 41(4). 1-23.

López, Belem G. & Vaid, Jyotsna. 2016. Divergence and overlap in bilingual conceptual representation: Does prior language brokering experience matter? *Bilingualism: Language and Cognition* 21(1). 1-12.

López, Belem G. & Vaid, Jyotsna. 20l8. Fácil or a piece of cake: Does variability in language brokering experience affect idiom comprehension? *Bilingualism: Language and Cognition* 21 (2). 340-354.

López, Belem G. & Vaid, Jyotsna & Tosun, Sümeyra & Rao, Chaitra. 2017. Bilinguals' plausibility judgements for phrases with a literal vs. non-literal meaning: The influence of language brokering experience. *Frontiers in Psychology* 8. 1-8.

McQuillan, Jeff & Tse, Lucy. 1995. Child language brokering in linguistic minority communities: Effects on cultural interaction, cognition, and literacy. *Language and Education* 9(3). 195-215.

Rainey, Vanessa R. & Davidson, Denise & Li-Grining, Christine. 2015. Executive functions as predictors of syntactic awareness in English monolingual and English–Spanish bilingual language brokers and non-brokers. *Applied Psycholinguistics* 37(4). 963-995.

Siyanova-Chanturia, Anna & Conklin, Kathy & Schmitt, Norbert. 2011. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27(2). 251-272.

Titone, Debra A. & Connine, Cynthia M. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31(12). 1655-1674.

Tse, Lucy. 1996. Language brokering in linguistic minority communities: The case of Chinese-and Vietnamese American students. *The Bilingual Research Journal* 20(3-4). 485-498.

Tzou, Yeh-Zu & Vaid, Jyotsna & Chen, Hsin-Chin. 2017. Does formal training in translation/interpreting affect translation strategy? Evidence from idiom translation. *Bilingualism: Language and Cognition* 20(3). 632-641.

Vaid, Jyotsna & Choi, Hyun & Chen, Hsin-Chin & Friedman, Michael. 2008. Perceiving and responding to embarrassing predicaments across language: Cultural influences on the emotion lexicon. *The Mental Lexicon* 3(1). 121-147.

Vaid, Jyotsna & López, Belem G. & Martinez, Francisco. 2015. Linking the figurative to the creative: Bilinguals' comprehension of metaphors, jokes, and remote associates. In Heredia, Roberto R. & Anna B. Cieslicka (eds.), *Bilingual figurative language processing,* 53-86. Cambridge: Cambridge University Press.

Weisskirch, Robert S. 2006. Emotional aspects of language brokering among Mexican American adults. *Journal of Multilingual and Multicultural Development* 27. 332-343.

Weisskirch, Robert S. 2010. Child language brokers in immigrant families: An overview of family dynamics. *mediAzioni* 10. 68-87.

Zhu, Zhu & Freeman, Helen. 2022. *Cambridge Dictionary of English – Chinese Translation.* Available online at https://dictionary.cambridge.org/dictionary/english-chinese-simplified/.

Zhu, Tianshu & Minda, Jean Paul. 2021. An investigation of idiom processing advantage using translated familiar idioms. *Canadian Journal of Experimental Psychology* 75(2). 162-168.

This article was first published at https://www.lacussquare.org